SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN

A Matter of TIME

It begins, it ends, it's real, it's an illusion. It's the ultimate paradox

ULTIMATE CLOCKS

THE PHILOSOPHY OF TIME

THE MIND AND TIME

TIME'S MYSTERIOUS PHYSICS

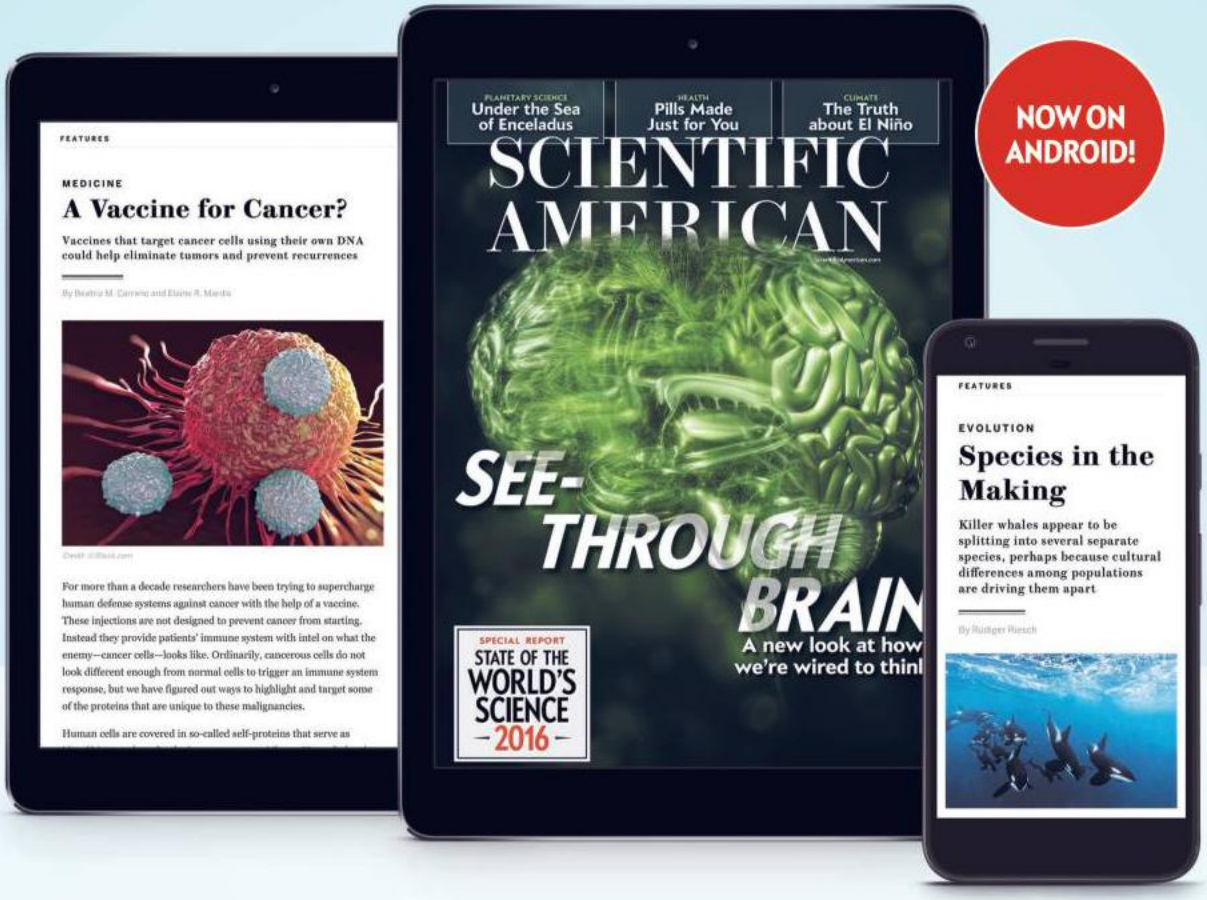
BUILDING TIME MACHINES

TIME AND CULTURE

SUMMER 2018
SCIENTIFIC AMERICAN SPECIALS

© 2018 Scientific American

Access *Scientific American* anytime, anywhere.


FEATURES
MEDICINE
A Vaccine for Cancer?
Vaccines that target cancer cells using their own DNA could help eliminate tumors and prevent recurrences
PLANETARY SCIENCE
Under the Sea of Enceladus
HEALTH
Pills Made Just for You
CLIMATE
The Truth about El Niño
SCIENTIFIC AMERICAN
SEE-THROUGH BRAIN
A new look at how we're wired to think
SPECIAL REPORT
STATE OF THE WORLD'S SCIENCE 2016
NOW ON ANDROID!
FEATURES
EVOLUTION
Species in the Making
Killer whales appear to be splitting into several separate species, perhaps because cultural differences among populations are driving them apart

Download the App for iOS™ and Android™

Download on the App Store

GET IT ON Google Play

Full Content
Read the latest issue on your phone or tablet—online or offline.

Easy Navigation
Browse articles and past issues quickly with thumbnail views.

Resize Text
Every article customized to your liking—on phone or tablet screens.

Scrapbook
Save your favorite articles to your scrapbook for future reference.

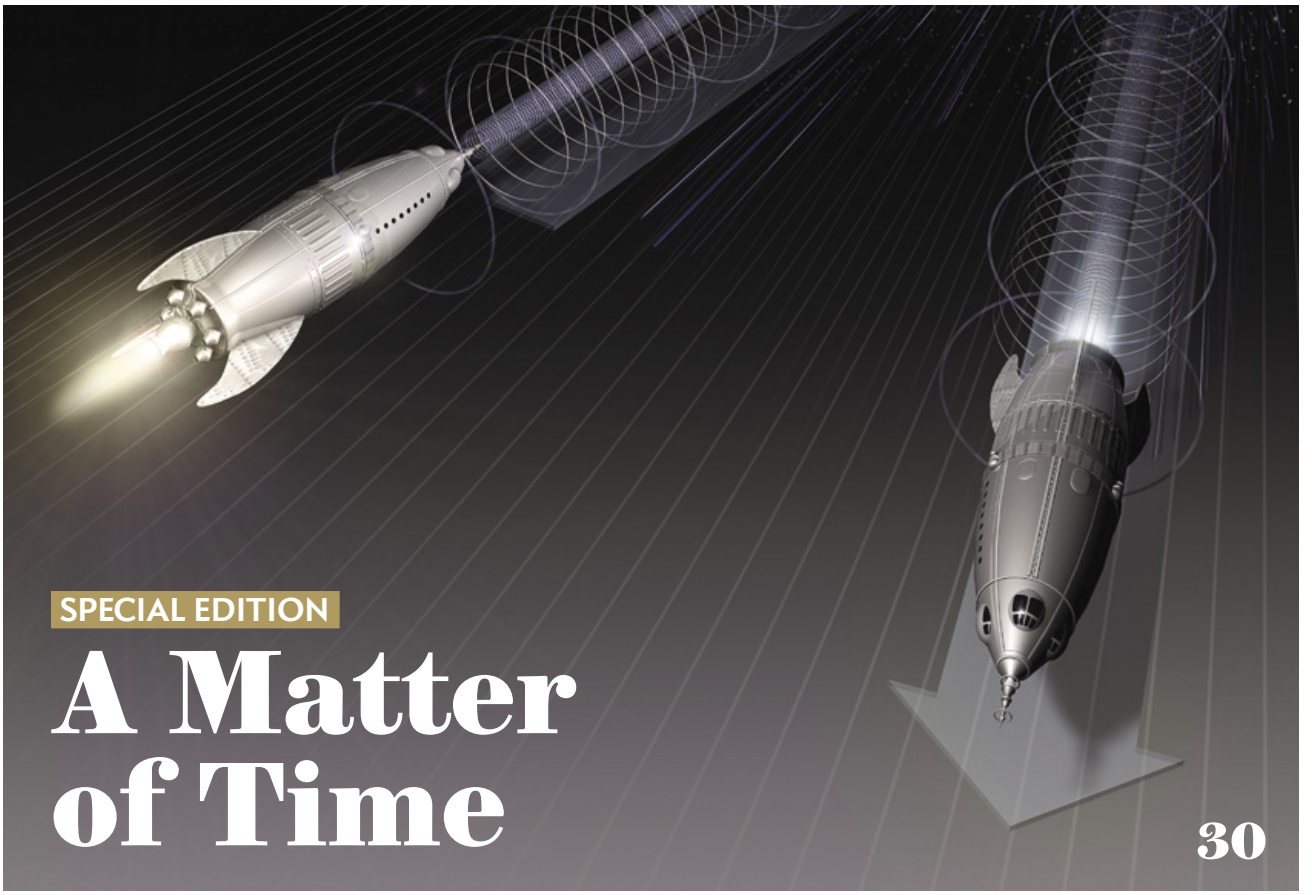
ARCHIVE

Done Reading?
Remove issues easily with the Archive button. You can re-download anytime.

Apple and the Apple logo are trademarks of Apple Inc., registered in the U.S. and other countries. App Store is a service mark of Apple Inc. Android and the Google Play logo are trademarks of Google Inc. IOS is a trademark or registered trademark of Cisco in the U.S. and other countries and is used under license.

SCIENTIFIC AMERICAN

Volume 27, Number 2, Summer 2018



SPECIAL EDITION

A Matter of Time

30

3 Introduction

HISTORY

4 Real Time
The pace of living quickens continuously, yet a full understanding of things temporal still eludes us. *By Gary Stix*

CHRONOLOGY

8 That Mysterious Flow
From the fixed past to the tangible present to the undecided future, it feels as though time flows inexorably on. But that is an illusion. *By Paul Davies*

PHYSICS

14 Is Time an Illusion?
The concepts of time and change may emerge from a universe that, at root, is utterly static. *By Craig Callender*

PHILOSOPHY

22 A Hole at the Heart of Physics
Physicists can't seem to find the time—literally. Can philosophers help? *By George Musser*

TRAVEL

24 How to Build a Time Machine
It wouldn't be easy, but it might be possible. *By Paul Davies*

THOUGHT EXPERIMENT

30 Time and the Twin Paradox
Does time tick by at the same rate for everyone? *By Ronald C. Lasky*

BIOLOGY

34 Times of Our Lives
Whether they're counting minutes, months or years, biological clocks help to keep our brains and bodies running on schedule. *By Karen Wright*

NEUROLOGY

42 Remembering When
Several brain structures contribute to "mind time," organizing our experiences into chronologies of remembered events. *By Antonio Damasio*

ANTHROPOLOGY

48 Clocking Cultures
What is time? The answer varies from society to society. *By the Editors*

COVER IMAGE BY TOM DRAPER DESIGN

Illustration by Malcolm Godwin

© 2018 Scientific American

INVENTION
50 A Chronicle of Timekeeping
Our conception of time depends on the way we measure it. *By William J. H. Andrewes*

NUMBERS
58 From Instantaneous to Eternal
The units of time range from the infinitesimally brief to the interminably long. The descriptions given here attempt to convey a sense of this vast chronological span. *By David Labrador*

ENGINEERING
60 Ultimate Clocks
Atomic clocks are shrinking to microchip size, heading for space—and approaching the limits of useful precision. *By W. Wayt Gibbs*

RELATIVITY
68 How Time Flies
Recent experimental optical clocks are so precise that even a small change in elevation or velocity makes them register the passage of time differently. *By John Matson*

ASTROPHYSICS
70 Inconstant Constants
Do the inner workings of nature change with time? *By John D. Barrow and John K. Webb*

COSMOLOGY
78 The Myth of the Beginning of Time
String theory suggests that the big bang was not the origin of the universe but simply the outcome of a preexisting state. *By Gabriele Veneziano*

INTERVIEW
90 What Keeps Time Moving Forward?
A timely conversation with physicist Sean M. Carroll about how our one-way trip from the past to the future is entangled with entropy and the origin of the universe. *By John Matson*

PHYSICS
94 Atoms of Space and Time
We perceive space and time to be continuous, but if the amazing theory of loop quantum gravity is correct, they actually come in discrete pieces. *By Lee Smolin*

COSMOLOGY
104 Could Time End?
Yes. And no. For time to end seems impossible and inevitable. Recent work in physics suggests a resolution to the paradox. *By George Musser*

END NOTE
112 Stopping Time
Stop-action strobe photography makes it possible to freeze time—and reveals phenomena too ephemeral to be seen with the naked eye.

Articles in this special issue are updated from previous issues of *Scientific American*.

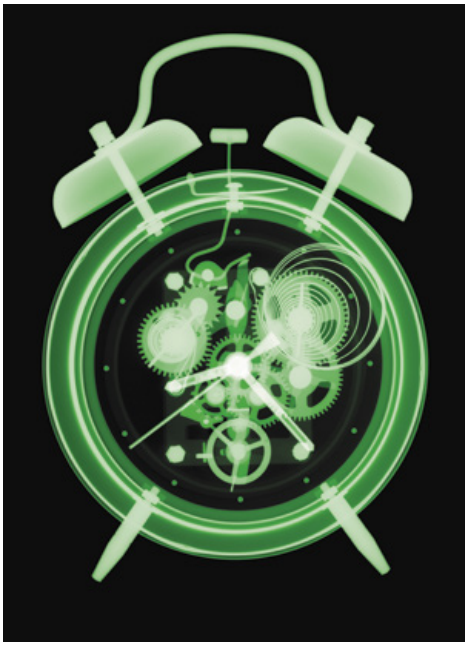
34 A "stopwatch" in your brain marks off time.

94 Like matter, time might be granular.

104 Time itself could fall apart in the future.

Copyright © 2018 Scientific American, a division of Nature America, Inc. All rights reserved. Scientific American Special (ISSN 1936-1513), Volume 27, Number 2, Summer 2018, published by Scientific American, a division of Nature America, Inc., 1 New York Plaza, Suite 4500, New York, NY 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, 1 New York Plaza, Suite 4500, New York, NY 10004-1562. Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.

© 2018 Scientific American



NICK VEASEY Getty Images

What Time Is It?

The chronic complaint

THAT SIMPLE QUESTION PROBABLY is asked more often today than ever. In our clock-studded, cellphone society, the answer is never more than a glance away, and so we can blissfully partition our days into ever smaller increments for ever more tightly scheduled tasks, confident that we will always know it is 7:03 P.M.

Modern scientific revelations about time, however, make the question endlessly frustrating. If we seek a precise knowledge of the time, the elusive infinitesimal of "now" dissolves into a scattering flock of nanoseconds. Bound by the speed of light and the velocity of nerve impulses, our perceptions of the present sketch the world as it was an instant ago—for all that our consciousness pretends otherwise, we can never catch up.

Even in principle, perfect synchronicity escapes us. Relativity dictates that, like a strange syrup, time flows slower on moving trains than in the stations and faster in the mountains than in the valleys. The time for our wristwatch or digital screen is not exactly the same as the time for our head. It is roughly 7:04 P.M.

Our intuitions are deeply paradoxical. Time heals all wounds, but it is also the great destroyer. Time is relative but also relentless. There is time for every purpose under heaven, but there is never enough. Time flies, crawls and races. Seconds can be both split and stretched. Like the tide, time waits for no man, but in dramatic moments it also stands still. It is as personal as the pace of one's heartbeat but as public as the clock tower in the town square. We do our best to reconcile the contradictions. It seems like 7:05 P.M.

And of course, time is money. It is the partner of change, the antagonist of speed, the currency in which we pay attention. It is our most precious, irreplaceable commodity. Yet still we say we don't know where it goes, and we sleep away a third of it, and none of us really can account for how much we have left. We can find 100 ways to save time, but the amount remaining nonetheless diminishes steadily. It is already 7:06 P.M.

Time and memory shape our perceptions of our own identity. We may feel ourselves to be at history's mercy, but we also see ourselves as free-willed agents of the future. That conception is disturbingly at odds with the ideas of physicists and philosophers, however, because if time is a dimension like those of space, then yesterday, today and tomorrow are all equally concrete and determined. The future exists as much as the past does; it is just in a place that we have not yet visited. Somewhere it is 7:07 P.M.

"Time is the substance of which I am made," Argentine writer Jorge Luis Borges wrote. "Time is a river that carries me away, but I am the river; it is a tiger that destroys me, but I am the tiger; it is a fire that consumes me, but I am the fire." This special issue of *Scientific American* summarizes what science has discovered about how time permeates and guides both our physical world and our inner selves. That knowledge should enrich the imagination and provide practical advantages to anyone hoping to beat the clock or at least to stay in step with it. It is now 7:08 P.M. Synchronize your watches. SA

THE EDITORS
editors@sciam.com

SCIENTIFIC AMERICAN®
ESTABLISHED 1845

A Matter of Time is published by the staff of SCIENTIFIC AMERICAN, with project management by:

Editor in Chief and Senior Vice President: **Mariette DiChristina**

Digital Content Manager: **Curtis Brainard**

Collections Editor: **Andrea Gawrylewski**

Creative Director: **Michael Mrak**
Issue Designer: **Caitlin Choi**
Senior Graphics Editor: **Jen Christiansen**
Photography Editor: **Monica Bradley**
Assistant Photo Editor: **Liz Tormes**

Copy Director: **Maria-Christina Keller**
Senior Copy Editor: **Daniel C. Schlenoff**
Copy Editor: **Aaron Shattuck**
Managing Production Editor: **Richard Hunt**
Prepress and Quality Manager: **Silvia De Santis**

Editorial Administrator: **Ericka Skirpan**
Senior Secretary: **Maya Harty**

Advertising Production Controller: **Carl Cherebin**
Production Controller: **Madelyn Keyes-Milch**

President: **Dean Sanderson**
Executive Vice President: **Michael Florek**
Executive Vice President, Global Advertising and Sponsorship: **Jack Laschever**
Publisher and Vice President: **Jeremy A. Abbate**
Associate Vice President, Business Development: **Diane McGarvey**
Head, Marketing and Product Management: **Richard Zinken**
Marketing Director, Institutional Partnerships and Customer Development: **Jessica Cole**
Online Marketing Product Manager: **Zoya Lysak**
Digital Marketing Manager: **Marie Cummings**
Marketing and Customer Service Coordinator: **Christine Kaelin**
Head, Communications, USA: **Rachel Scheer**
Director, Integrated Media: **Jay Berfas**
Director, Integrated Media: **Matt Bondlow**
Director, Global Media Alliances: **Ted Macauley**
Custom Publishing Editor: **Lisa Pallatroni**
Rights and Permissions Manager: **Felicia Ruocco**
Senior Administrator, Executive Services: **May Jung**

© 2018 Scientific American

© 2018 Scientific American

DANIEL VFUNG Getty Images

Gary Stix is a senior editor at *Scientific American*.

HISTORY

REAL TIME

The pace of living quickens continuously, yet a full understanding of things temporal still eludes us

By Gary Stix

MORE THAN 200 YEARS AGO BENJAMIN Franklin coined the now famous dictum that equated passing minutes and hours with shillings and pounds. The new millennium—and the decades leading up to it—has given his words their real meaning. Time has become to the 21st century what fossil fuels and precious metals were to previous epochs. Constantly measured and priced, this vital raw material continues to spur the growth of economies built on a foundation of terabytes and gigabits per second.

© 2018 Scientific American

An English economics professor even tried to capture the millennial zeitgeist by supplying Franklin's adage with a quantitative underpinning. According to a formula derived by Ian Walker, now at Lancaster University Management School, three minutes of brushing one's teeth works out to the equivalent of 49 cents, the compensation (after taxes and Social Security) that the average Briton gives up by doing something besides working. Half an hour of washing a car by hand translates into $4.90.

This reduction of time to money may extend Franklin's observation to an absurd extreme. Still, the commodification of time is genuine—and results from a radical alteration in how we view the passage of events. Our fundamental human drives have not changed from the Paleolithic era, hundreds of thousands of years ago. Much of what we are about centers on the same impulses to eat, procreate, fight or flee that motivated Fred Flintstone. Despite the constancy of these primal urges, human culture has experienced upheaval after upheaval in the period since our hunter-gatherer forebears roamed the savannas. Perhaps the most profound change in the long transition from Stone Age to Information Age revolves around our subjective experience of time.

By one definition, time is a continuum in which one event follows another from the past through to the future. Today the number of occurrences packed inside a given interval, whether it be a year or a nanosecond, increases unendingly. The technological age has become a game of one-upmanship in which more is always better. In his 2000 book *Faster: The Acceleration of Just About Everything,* James Gleick notes that before Federal Express shipping became commonplace in the 1980s, the exchange of business documents did not usually require a package to be delivered "absolutely positively overnight." At first, FedEx gave its customers an edge. Soon, though, the whole world expected goods to arrive the next morning. "When everyone adopted overnight mail, equality was restored," Gleick writes, "and only the universally faster pace remained."

SIMULTANEITY

THE ADVENT of the Internet eliminated the burden of having to wait until the next day for the FedEx truck. In Internet time, everything happens everywhere at once—connected computer users can witness an update to a Web page at an identical moment in New York City or Dakar. Time has, in essence, triumphed over space. Noting this trend, Swatch, the watchmaker, went so far as to try to abolish the temporal boundaries that separate one place from another. It created a standard for Internet timekeeping that eliminated time zones, dividing the day into 1,000 increments that are the same anywhere on the globe, with the meridian at Biel, Switzerland, the location of Swatch's headquarters.

You can still put a digital Internet clock through its paces on the Web. But the prospects for it as a widely adopted universal time standard are about as good as the frustrated aspirations for Esperanto to become the world's lingua franca.

Leaving gimmickry aside, the wired world does erase time

IN INTERNET TIME, EVERYTHING HAPPENS EVERYWHERE AT ONCE. TIME HAS TRIUMPHED OVER SPACE.

barriers. This achievement relies on an ever progressing ability to measure time more precisely. Over the aeons, the capacity to gauge duration has correlated directly with increasing control over the environment that we inhabit. Keeping time is a practice that may go back more than 20,000 years, when hunters of the Ice Age notched holes in sticks or bones, possibly to track the days between phases of the moon. A mere 5,000 years ago or so the Babylonians and Egyptians devised calendars for planting and other time-sensitive activities.

Early chronotechnologists were not precision freaks. They tracked natural cycles: the solar day, the lunar month and the solar year. The sundial could do little more than cast a shadow, when clouds or night did not render it a useless decoration. Beginning in the 13th century, though, the mechanical clock initiated a revolution equivalent to the one engendered by the later invention by Gutenberg of the printing press. Time no longer "flowed," as it did literally in a water clock. Rather it was marked off by a mechanism that could track the beats of an oscillator. When refined, this device let time's passage be counted to fractions of a second.

The mechanical clock ultimately enabled the miniaturization of the timepiece. Once it was driven by a coiled spring and not a falling weight, it could be carried or worn like jewelry. The technology changed our perception of the way society was organized. It was an instrument that let one person coordinate activities with another. "Punctuality comes from within, not from without," wrote Harvard University historian David S. Landes in his 2000 book *Revolution in Time.* "It is the mechanical clock that made possible, for better or worse, a civilization attentive to the passage of time, hence to productivity and performance."

Mechanical clocks persisted as the most accurate timekeepers for centuries. But the past 50 years have seen as much progress in the quest for precision as in the previous 700 [see "A Chronicle of Timekeeping," by William J. H. Andrewes, on page 50]. It hasn't been just the Internet that has brought about the conquest of time over space. Time is more accurately measured than any other physical entity. As such, elapsed time is marshaled to size up spatial dimensions. Today standard makers gauge the length of the venerable meter by the distance light in a vacuum travels in 1/299,792,458th of a second.

Atomic clocks, which are used to make such measurements,

IN BRIEF

In the 21st century time has become a precious commodity, not only measured but also priced.

Keeping time is a practice that may go back more than 20,000 years, but it has become increasingly precise. Today time is more accurately measured than any other physical entity.

One of the hottest questions is whether time itself is illusory. This puzzle preoccupies not only physicists but also philosophers and anthropologists.

Ultimately it does not matter whether it really exists. Our reverence for time rests on a deep psychic need to recognize meaningful milestones.

© 2018 Scientific American

also play a role in judging location. In some of them, the resonant frequency of cesium atoms remains amazingly stable, becoming a pseudo pendulum capable of maintaining near nanosecond precision. Global Positioning System (GPS) satellites continuously broadcast their exact whereabouts as well as the time maintained by onboard atomic clocks. A receiving device processes this information from at least four satellites into exact terrestrial coordinates for the pilot or the hiker, whether in Patagonia or Lapland. The requirements are exacting. A time error of only a millionth of a second from an individual satellite could send a signal to a GPS receiver that would be inaccurate by as much as a fifth of a mile (if it went uncorrected by other satellites).

Advances in precision timekeeping continue apace. In fact, clockmakers may soon outdo themselves. They may create an atomic clock so precise that it will be impossible to synchronize other timepieces to it [see "Ultimate Clocks," by W. Wayt Gibbs, on page 60]. Researchers also continue to press ahead in slicing and dicing the second more finely. The need for speed has become a cornerstone of the Information Age. The world's speediest transistors can switch faster than a picosecond, a thousandth of a billionth of a second [see "From Instantaneous to Eternal," on page 58].

A team from France and the Netherlands set a speed record for subdividing the second, reporting in 2001 that a laser strobe light had emitted pulses lasting 250 attoseconds—that is 250 billionths of a billionth of a second—and this milestone has since been bested. The strobe may one day be fashioned into a camera that can track the movements of single electrons. The modern era has also registered gains in assessing big intervals. Radiometric dating methods, measuring rods of "deep time," indicate how old the earth really is.

The ability to transcend time and space effortlessly—whether on the Internet or piloting a GPS-guided airliner—lets us do things faster. Just how far speed limits can be stretched remains to be tested. Conference sessions and popular books toy with ideas for the ultimate cosmic hot rod, a means of traveling forward or back in time [see "How to Build a Time Machine," by Paul Davies, on page 24]. Yet despite watchmakers' prowess, neither physicists nor philosophers have come to any agreement about what we mean when we say "tempus fugit."

Perplexity about the nature of time—a tripartite oddity that parses into past, present and future—precedes the industrial era by quite a few centuries. Augustine described the definitional dilemma more eloquently than anyone. "What, then, is time?" he asked in his *Confessions*. "If no one asks me, I know; if I want to explain it to someone who does ask me, I do not know." He then went on to attempt to articulate why temporality is so hard to define: "How, then, can these two kinds of time, the past and the future, be, when the past no longer is and the future as yet does not be?"

Hard-boiled physicists, unburdened by theistic encumbrances, have also had difficulty grappling with this question. We remark that time "flies" as we hurtle toward our inevitable demise. What does that mean exactly? Saying that time races along at one second per second has as much scientific weight as the utterance of a Zen koan. One could hypothesize a metric of current flow for time, a form of temporal amperage. But such a measure may simply not exist [see "That Mysterious Flow," by Paul Davies, on page 8]. In fact, one of the hottest themes in theoretical physics is whether time itself is illusory. The confusion is such that physicists have gone as far as to recruit philosophers in their attempt to understand whether a *t* variable should be added to their equations [see "A Hole at the Heart of Physics," by George Musser, on page 22].

THE GREAT MANDALA

THE ESSENCE OF TIME is an age-old conundrum that preoccupies not just the physicist and the philosopher but also the anthropologist who studies non-Western cultures that perceive events as proceeding in a cyclical, nonlinear sequence [see "Clocking Cultures," by the Editors, on page 48]. Yet for most of us, time is not only real, it is the master of everything we do.

The distinct feeling we have of being bookended between a past and a future—or, in a traditional culture, being enmeshed in the Great Mandala of recurring natural rhythms—may be related to a basic biological reality. Our bodies are chock-full of living clocks—ones that govern how we connect a ball with a bat, when we feel sleepy and perhaps even when our time is up [see "Times of Our Lives," by Karen Wright, on page 34].

These real biorhythms have now begun to reveal themselves to biologists. Scientists are closing in on areas of the brain that produce the sensation of time flying when we're having fun—the same places that induce the slow-paced torpor of sitting through a monotone lecture on Canadian interest-rate policy. They are also beginning to understand the connections between different kinds of memory and how events are organized and recalled chronologically. Studies of neurological patients with various forms of amnesia, some of whom have lost the ability to judge accurately the passage of hours, months and even entire decades, are helping to pinpoint which areas of the brain are involved in how we experience time [see "Remembering When," by Antonio Damasio, on page 42].

Recalling where we fit in the order of things determines who we are. Ultimately it doesn't matter whether time, in cosmological terms, retains an underlying physical truth. If it is a fantasy, it is one we cling to steadfastly. The reverence we hold for the fourth dimension, the complement of the three spatial ones, has much to do with a deep psychic need to embrace meaningful temporal milestones that we can all share: birthdays, Christmas, the Fourth of July. How else to explain the frenzy of celebration in January 2000 for a date that neither marked a highlight of Christ's life nor, by many tallies, the true millennium?

We will, nonetheless, continue to celebrate the next millennium (if we as a species are still around), and in the meantime we will fete our parents' golden wedding anniversary and the 20th year of the founding of our local volunteer fire department. Doing so seems to be the only way of imposing hierarchy and structure on a world in which text messaging, express checkout and same-day delivery threaten to rob us of any sense of permanence.

MORE TO EXPLORE

The Story of Time. Edited by Kristen Lippincott. Merrell Holberton, 1999.
Faster: The Acceleration of Just About Everything. James Gleick. Vintage, 2000.
Revolution in Time: Clocks and the Making of the Modern World. Revised edition. David S. Landes. Belknap Press, 2000.
The Discovery of Time. Edited by Stuart McCready. Sourcebooks, 2001.

scientificamerican.com/magazine/sa

© 2018 Scientific American

Paul Davies is a theoretical physicist and cosmologist at Arizona State University, where he is director of BEYOND: Center for Fundamental Concepts in Science. He is one of the most prolific writers of popular-level books in physics.

CHRONOLOGY

THAT MYSTERIOUS FLOW

From the fixed past to the tangible present to the undecided future, it feels as though time flows inexorably on. But that is an illusion

By Paul Davies

IN BRIEF

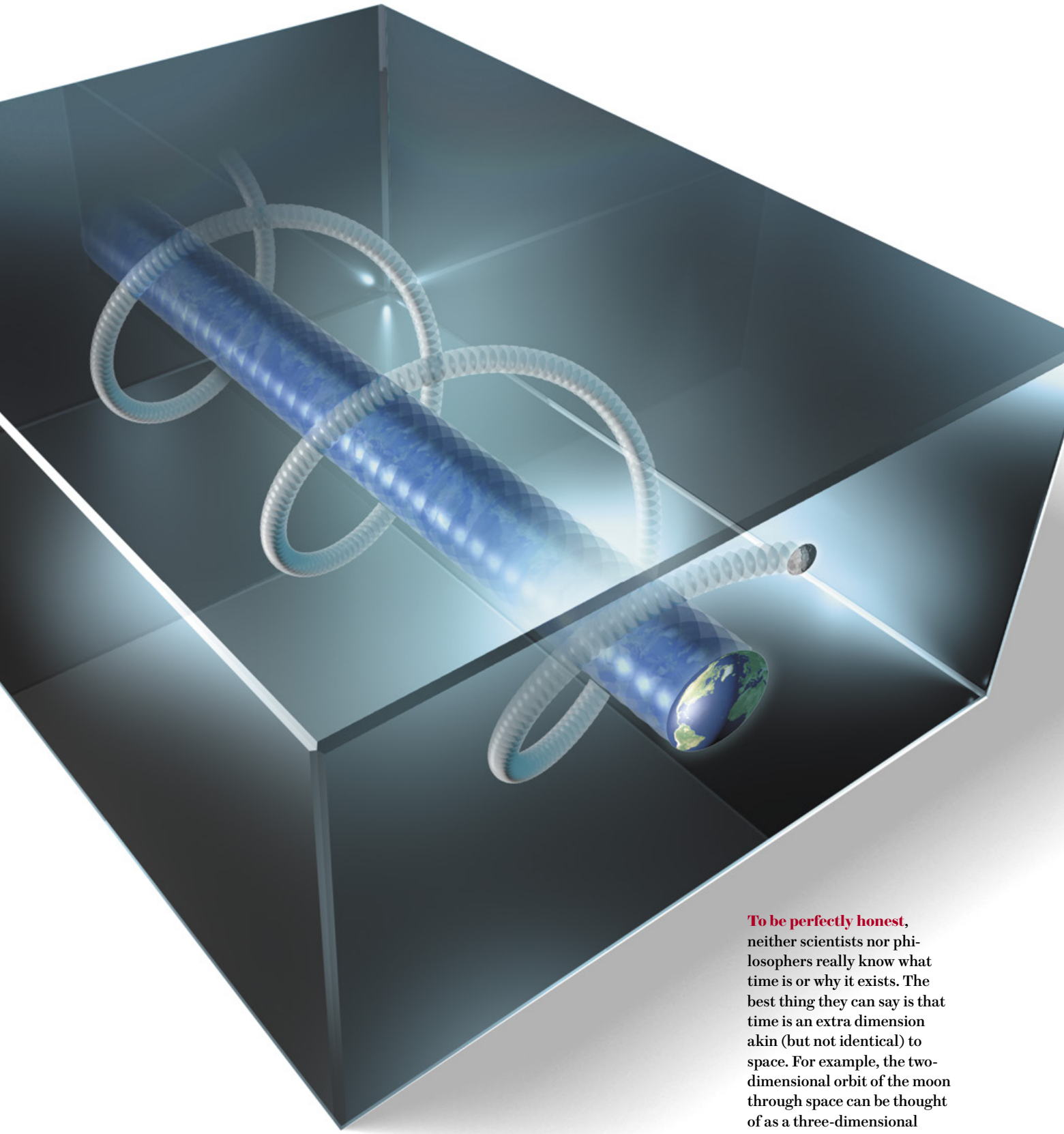
Our senses tell us that time flows: namely, that the past is fixed, the future is undetermined, and reality lives in the present. Yet various physical and philosophical arguments suggest otherwise.

The passage of time is most likely an illusion. Consciousness may involve thermodynamic or quantum processes that lend the impression of living moment by moment.

"GATHER YE ROSEBUDS WHILE YE MAY,/OLD TIME IS STILL A-FLYING." So wrote 17th-century English poet Robert Herrick, capturing the universal cliché that time flies. And who could doubt that it does? The passage of time is probably the most basic facet of human perception, for we feel time slipping by in our innermost selves in a manner that is altogether more intimate than our experience of, say, space or mass. The passage of time has been compared to the flight of an arrow and to an ever rolling stream, bearing us inexorably from past to future. Shakespeare wrote of "the whirligig of time," his countryman Andrew Marvell of "Time's winged chariot hurrying near."

Evocative though these images may be, they run afoul of a deep and devastating paradox. Nothing in known physics corresponds to the passage of time. Indeed, physicists insist that time doesn't flow at all; it merely is. Some philosophers argue that the very notion of the passage of time is nonsensical and that talk of the river or flux of time is founded on a misconception. How can something so basic to our experience of the physical world turn out to be a case of mistaken identity? Or is there a key quality of time that science has not yet identified?

© 2018 Scientific American



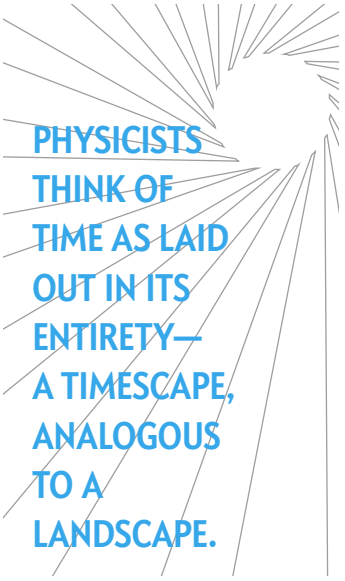
To be perfectly honest, neither scientists nor philosophers really know what time is or why it exists. The best thing they can say is that time is an extra dimension akin (but not identical) to space. For example, the two-dimensional orbit of the moon through space can be thought of as a three-dimensional corkscrew through spacetime.

Illustrations by Bryan Christie Design

© 2018 Scientific American

TIME ISN'T OF THE ESSENCE

IN DAILY LIFE we divide time into three parts: past, present and future. The grammatical structure of language revolves around this fundamental distinction. Reality is associated with the present moment. The past we think of as having slipped out of existence, whereas the future is even more shadowy, its details still unformed. In this simple picture, the "now" of our conscious awareness glides steadily onward, transforming events that were once in the unformed future into the concrete but fleeting reality of the present—and thence relegating them to the fixed past.



Obvious though this commonsense description may seem, it is seriously at odds with modern physics. Albert Einstein famously expressed this point when he wrote to a friend, "The past, present and future are only illusions, even if stubborn ones." Einstein's startling conclusion stems directly from his special theory of relativity, which denies any absolute, universal significance to the present moment. According to the theory, simultaneity is relative. Two events that occur at the same moment if observed from one reference frame may occur at different moments if viewed from another.

An innocuous question such as "What is happening on Mars now?" has no definite answer. The key point is that Earth and Mars are a long way apart—up to about 20 light-minutes. Because information cannot travel faster than light, an Earth-based observer is unable to know the situation on Mars at the same instant. He must infer the answer after the event, when light has had a chance to pass between the planets. The inferred past event will be different depending on the observer's velocity.

For example, during a future manned expedition to Mars, mission controllers back on Earth might say, "I wonder what Commander Jones is doing at Alpha Base now." Looking at their clock and seeing that it was 12:00 P.M. on Mars, their answer might be "eating lunch." But an astronaut zooming past Earth at near the speed of light at the same moment could, on looking at his clock, say that the time on Mars was earlier or later than 12:00 P.M., depending on his direction of motion. That astronaut's answer to the question about Commander Jones's activities would be "cooking lunch" or "washing dishes" [*see box on page 12*]. Such mismatches make a mockery of any attempt to confer special status on the present moment, for whose "now" does that moment refer to? If you and I were in relative motion, an event that I might judge to be in the as yet undecided future might for you already exist in the fixed past.

The most straightforward conclusion is that both past and future are fixed. For this reason, physicists prefer to think of time as laid out in its entirety—a timescape, analogous to a landscape—with all past and future events located there together. It is a notion sometimes referred to as block time. Completely absent from this description of nature is anything that singles out a privileged, special moment as the present or any process that would systematically turn future events into present, then past, events. In short, the time of the physicist does not pass or flow.

HOW TIME DOESN'T FLY

A NUMBER OF PHILOSOPHERS over the years have arrived at the same conclusion by examining what we normally mean by the passage of time. They argue that the notion is internally inconsistent. The concept of flux, after all, refers to motion. It makes sense to talk about the movement of a physical object, such as an arrow through space, by gauging how its location varies with time. But what meaning can be attached to the movement of time itself? Relative to what does it move? Whereas other types of motion relate one physical process to another, the putative flow of time relates time to itself. Posing the simple question "How fast does time pass?" exposes the absurdity of the very idea. The trivial answer "one second per second" tells us nothing at all.

Although we find it convenient to refer to time's passage in everyday affairs, the notion imparts no new information that cannot be conveyed without it. Consider the following scenario: Alice was hoping for a white Christmas, but when the day came she was disappointed that it only rained; however, she was happy that it snowed the following day. Although this description is replete with tenses and references to time's passage, exactly the same information is conveyed by simply correlating Alice's mental states with dates, in a manner that omits all reference to time passing or the world changing. Thus, the following cumbersome and rather dry catalogue of facts suffices:

NOBODY REALLY KNOWS ...

What Is Time, Anyway?

Augustine of Hippo, the famous fifth-century theologian, remarked that he knew well what time is—until somebody asked. Then he was at a loss for words. Because we sense time psychologically, definitions of time based on physics seem dry and inadequate. For the physicist, time is simply what (accurate) clocks measure. Mathematically, it is a one-dimensional space, usually assumed to be continuous, although it might be quantized into discrete "chronons," like frames of a movie.

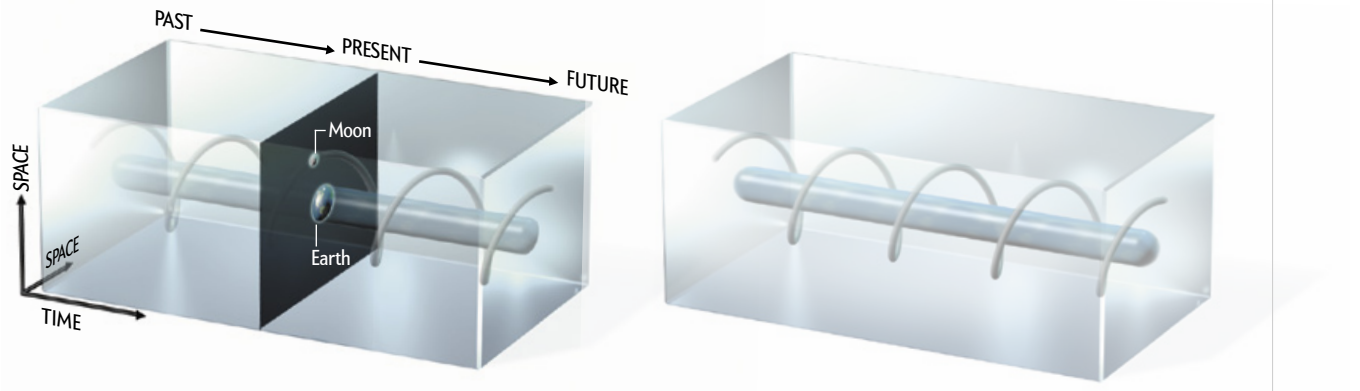
The fact that time may be treated as a fourth dimension does not mean that it is identical to the three dimensions of space. Time and space enter into daily experience and physical theory in distinct ways. For instance, the formula for calculating spacetime distances is not the same as the one for calculating spatial distances. The distinction between space and time underpins the key notion of causality, stopping cause and effect from being hopelessly jumbled. On the other hand, many physicists believe that on the very smallest scale of size and duration, space and time might lose their separate identities. *—P.D.*

© 2018 Scientific American

All Time Like the Present

According to conventional wisdom, the present moment has special significance. It is all that is real. As the clock ticks, the moment passes and another comes into existence—a process that we call the flow of time. The moon, for example, is located at only one position in its orbit around Earth. Over time it ceases to exist at that position and is instead found at a new position.

Researchers who think about such things, however, generally argue that we cannot single out a present moment as special when every moment considers itself to be special. Objectively, past, present and future must be equally real. All of eternity is laid out in a four-dimensional block composed of time and the three spatial dimensions. (This diagram shows only two of these spatial dimensions.)



Conventional view: Only the present is real

Block universe: All times are equally real

December 24: Alice hopes for a white Christmas.
December 25: There is rain. Alice is disappointed.
December 26: There is snow. Alice is happy.

In this description, nothing happens or changes. There are simply states of the world at different dates and associated mental states for Alice.

Similar arguments go back to ancient Greek philosophers such as Parmenides and Zeno. A century ago British philosopher John McTaggart sought to draw a clear distinction between the description of the world in terms of events happening, which he called the A series, and the description in terms of dates correlated with states of the world, the B series. Each appears to be a true description of reality, and yet the two points of view are seemingly in contradiction. For example, the event "Alice is disappointed" was once in the future, then in the present and afterward in the past. But past, present and future are exclusive categories, so how can a single event have the character of belonging to all three? McTaggart used this clash between the A and B series to argue for the unreality of time as such, perhaps a rather drastic conclusion. Most physicists would put it less dramatically: the flow of time is unreal, but time itself is as real as space.

JUST IN TIME

A GREAT SOURCE OF CONFUSION in discussions of time's passage stems from its link with the so-called arrow of time. To deny that time flows is not to claim that the designations "past" and "future" are without physical basis. Events in the world undeniably form a unidirectional sequence. For instance, an egg dropped on the floor will smash into pieces, whereas the reverse process—a broken egg spontaneously assembling itself into an intact egg—is never witnessed. This is an example of the second law of thermodynamics, which states that the entropy of a closed system—roughly defined as how disordered it is—will tend to rise with time. An intact egg has lower entropy than a shattered one.

Because nature abounds with irreversible physical processes, the second law of thermodynamics plays a key role in imprinting on the world a conspicuous asymmetry between past and future directions along the time axis. By convention, the arrow of time points toward the future. This does not imply, however, that the arrow is moving toward the future, any more than a compass needle pointing north indicates that the compass is traveling north. Both arrows symbolize an asymmetry, not a movement. The arrow of time denotes an asymmetry of the world in time, not an asymmetry or flux of time. The labels "past" and "future" may legitimately be applied to temporal directions, just as "up" and "down" may be applied to spatial directions, but talk of the past or the future is as meaningless as referring to the up or the down.

The distinction between pastness or futureness and "the" past or "the" future is graphically illustrated by imagining a movie of, say, the egg being dropped on the floor and breaking. If the film were run backward through the projector, everyone would see that the sequence was unreal. Now imagine if the film strip were cut up into frames and the frames shuffled randomly. It would be a straightforward task for someone to rearrange the stack of frames into a correctly ordered sequence, with the broken egg at the top of the stack and the intact egg at the bottom. This vertical stack retains the asymmetry implied by the arrow of time because it forms an ordered sequence in vertical space, proving that time's asymmetry is actually a property of states of the world, not a property of time as such. It is

© 2018 Scientific American

It's All Relative

What is happening on Mars right now? Such a simple question, such a complex answer. The trouble stems from the phrase "right now." Different people, moving at different velocities, have different perceptions of what the present moment is. This strange fact is known as the relativity of simultaneity. In the following scenario, two people—an earthling sitting in Houston and a rocket man crossing the solar system at 80 percent of the speed of light—attempt to answer the question of what is happening on Mars right now. A resident of Mars has agreed to eat lunch when his clock strikes 12:00 P.M. and to transmit a signal at the same time.

As Seen from Earth

From the earthling's perspective, Earth is standing still, Mars is a constant distance (20 light-minutes) away, and the rocket ship is moving at 80 percent of the speed of light. The situation looks exactly the same to the Martian.

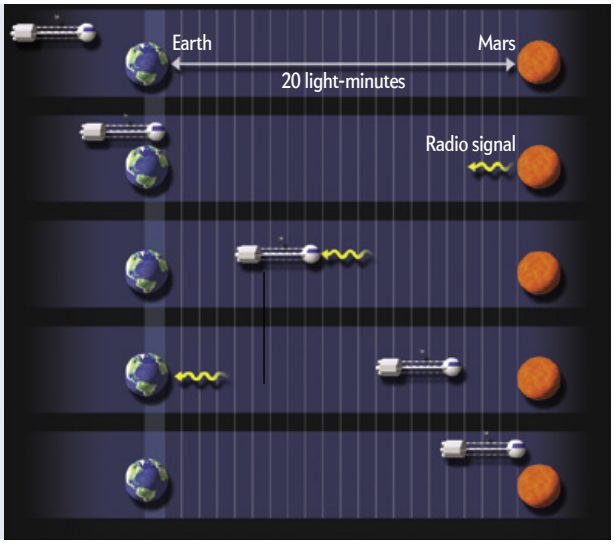
BEFORE NOON By exchanging light signals, the earthling and Martian measure the distance between them and synchronize their clocks.

12:00 P.M. The earthling hypothesizes that the Martian has begun to eat lunch. He prepares to wait 20 minutes for verification.

12:11 P.M. Knowing the rocket's speed, the earthling deduces that it encounters the signal while on its way to Mars.

12:20 P.M. The signal arrives at Earth. The earthling has confirmed his earlier hypothesis. Noon on Mars is the same as noon on Earth.

12:25 P.M. The ship arrives at Mars.



As Seen from the Rocket

From the rocket man's perspective, the rocket is standing still. It is the planets that are hurtling through space at 80 percent of the speed of light. His measurements show the two planets to be separated by 12 light-minutes—a different distance than the earthling inferred. This discrepancy, a well-known effect of Einstein's theory, is called length contraction. A related effect, time dilation, causes clocks on the ship and planets to run at different rates. (The earthling and Martian think the ship's clock is slow; the rocket man thinks the planets' are.) As the ship passes Earth, it synchronizes its clock to Earth's.

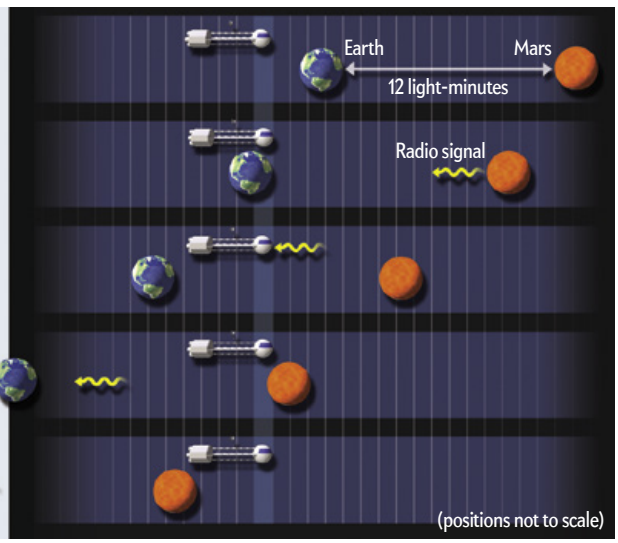
BEFORE NOON By exchanging light signals with his colleagues, the rocket man measures the distance between the planets.

12:00 P.M. Passing Earth, the rocket man hypothesizes that the Martian has begun to eat. He prepares to wait 12 minutes for verification.

12:07 P.M. The signal arrives, disproving the hypothesis. The rocket man infers that the Martian ate sometime before noon (rocket time).

12:15 P.M. Mars arrives at the ship. The rocket man and Martian notice that their two clocks are out of sync but disagree as to whose is right.

12:33 P.M. The signal arrives at Earth. The clock discrepancies demonstrate that there is no universal present moment.



© 2018 Scientific American

not necessary for the film to actually be run as a movie for the arrow of time to be discerned.

Given that most physical and philosophical analyses of time fail to uncover any sign of a temporal flow, we are left with something of a mystery. To what should we attribute the powerful, universal impression that the world is in a continual state of flux? Some researchers, notably the late Nobel laureate chemist Ilya Prigogine, have contended that the subtle physics of irreversible processes make the flow of time an objective aspect of the world. But I and others argue that it is some sort of illusion.

After all, we do not really observe the passage of time. What we actually observe is that later states of the world differ from earlier states that we still remember. The fact that we remember the past, rather than the future, is an observation not of the passage of time but of the asymmetry of time. Nothing other than a conscious observer registers the flow of time. A clock measures durations between events much as a measuring tape measures distances between places; it does not measure the "speed" with which one moment succeeds another. Therefore, it appears that the flow of time is subjective, not objective.

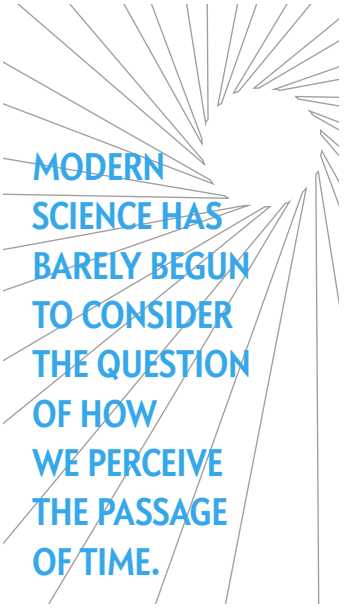
LIVING IN THE PRESENT

THIS ILLUSION CRIES OUT for explanation, and that explanation is to be sought in psychology, neurophysiology, and maybe linguistics or culture. Modern science has barely begun to consider the question of how we perceive the passage of time; we can only speculate about the answer. It might have something to do with the functioning of the brain. If you spin around several times and stop suddenly, you will feel giddy. Subjectively, it seems as if the world is rotating relative to you, but the evidence of your eyes is clear enough: it is not. The apparent movement of your surroundings is an illusion created by the rotation of fluid in the inner ear. Perhaps temporal flux is similar.

There are two aspects to time asymmetry that might create the false impression that time is flowing. The first is the thermodynamic distinction between past and future. As physicists have realized over the past few decades, the concept of entropy is closely related to the information content of a system. For this reason, the formation of memory is a unidirectional process—new memories add information and raise the entropy of the brain. We might perceive this unidirectionality as the flow of time.

A second possibility is that our perception of the flow of time is linked in some way to quantum mechanics. It was appreciated from the earliest days of the formulation of quantum mechanics that time enters into the theory in a unique manner, quite unlike space. The special role of time is one reason it is proving so difficult to merge quantum mechanics with general relativity. The Heisenberg uncertainty principle, according to which nature is inherently indeterministic, implies an open future (and, for that matter, an open past). This indeterminism manifests itself most conspicuously on an atomic scale of size and dictates that the observable properties that characterize a physical system are generally undecided from one moment to the next.

For example, an electron hitting an atom may bounce off in one of many directions, and it is normally impossible to predict in advance what the outcome in any given case will be. Quantum indeterminism implies that for a particular quantum state there are many (possibly infinite) alternative futures or potential realities. Quantum mechanics supplies the relative probabilities for each observable outcome, although it won't say which potential future is destined for reality.



But when a human observer makes a measurement, one and only one result is obtained; for example, the rebounding electron will be found moving in a certain direction. In the act of measurement, a single, specific reality gets projected out from a vast array of possibilities. Within the observer's mind, the possible makes a transition to the actual, the open future to the fixed past—which is precisely what we mean by the flux of time.

There is no agreement among physicists on how this transition from many potential realities into a single actuality takes place. Many physicists have argued that it has something to do with the consciousness of the observer, on the basis that it is the act of observation that prompts nature to make up its mind. A few researchers, such as Roger Penrose of the University of Oxford, maintain that consciousness—including the impression of temporal flux—could be related to quantum processes that take place in the brain.

Although researchers have failed to find evidence for a single "time organ" in the brain, in the manner of, say, the visual cortex, it may be that future work will pin down those brain processes responsible for our sense of temporal passage. It is possible to imagine drugs that could suspend the subject's impression that time is passing. Indeed, some practitioners of meditation claim to be able to achieve such mental states naturally.

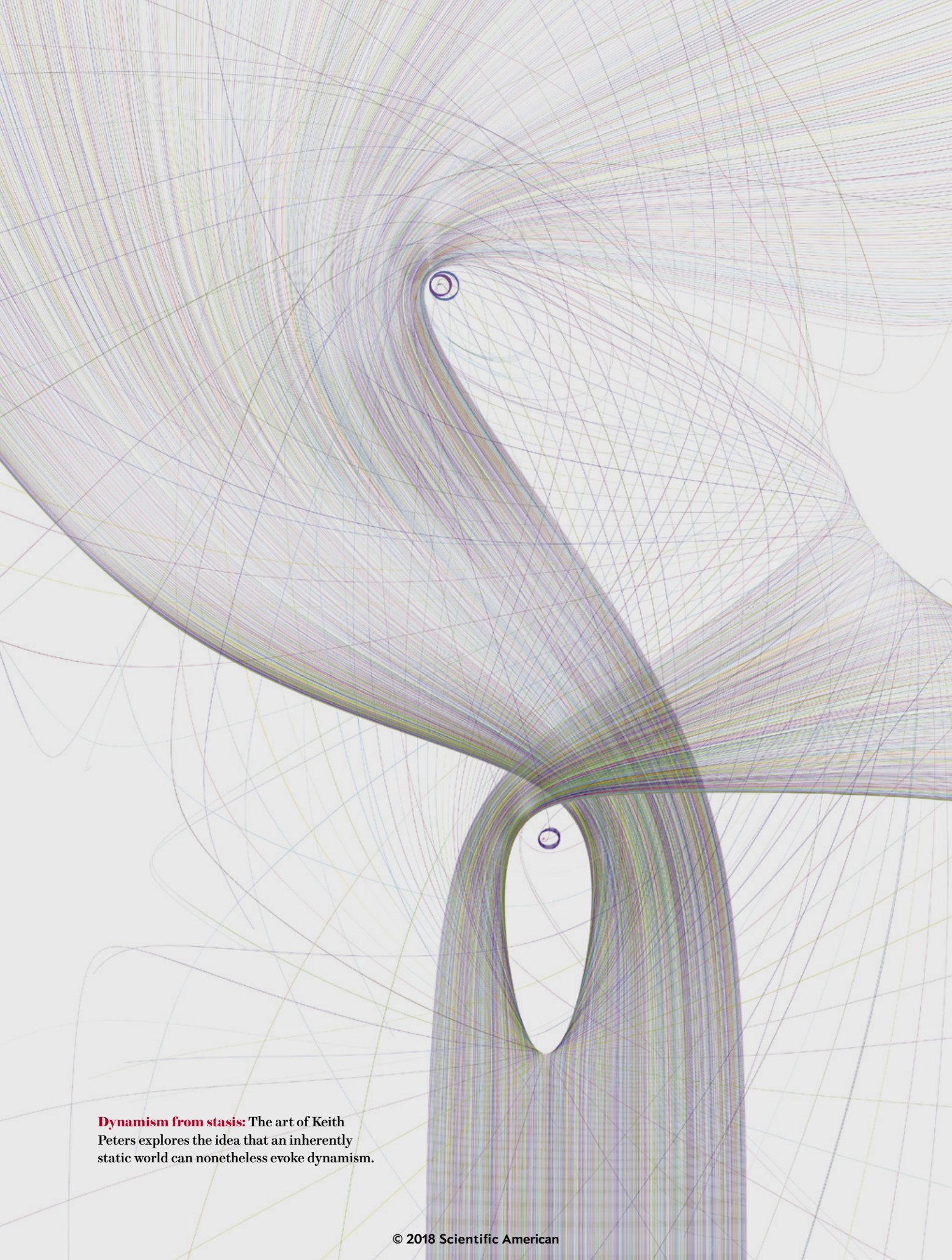
And what if science were able to explain away the flow of time? Perhaps we would no longer fret about the future or grieve for the past. Worries about death might become as irrelevant as worries about birth. Expectation and nostalgia might cease to be part of human vocabulary. Above all, the sense of urgency that attaches to so much of human activity might evaporate. No longer would we be slaves to Henry Wadsworth Longfellow's entreaty to "act, act in the living present," for the past, present and future would literally be things of the past. SA

MORE TO EXPLORE

The Unreality of Time. John Ellis McTaggart in *Mind,* Vol. 17, pages 456–473; 1908.
Can Time Go Backward? Martin Gardner in *Scientific American,* Vol. 216, No. 1, pages 98–108; January 1967.
The Physics of Time Asymmetry. Paul Davies. University of California Press, 1977.
Time and Becoming. J.J.C. Smart in *Time and Cause.* Edited by Peter van Inwagen. D. Reidel Publishing, 1980.
About Time: Einstein's Unfinished Revolution. Paul Davies. Simon & Schuster, 1996.
What Is Time? G. J. Whitrow. Oxford University Press, 2004.

scientificamerican.com/magazine/sa

© 2018 Scientific American



Dynamism from stasis: The art of Keith Peters explores the idea that an inherently static world can nonetheless evoke dynamism.

© 2018 Scientific American

Craig Callender is a philosophy professor at the University of California, San Diego. He wrote the graphic text *Introducing Time* (2001) and edited *The Oxford Handbook of Philosophy of Time* (2011). He assures us that his lifelong interest in time has nothing to do with his last name.

PHYSICS

IS TIME AN ILLUSION?

The concepts of time and change may emerge from a universe that, at root, is utterly static

By Craig Callender

As you read this sentence, you probably think that this moment—right now—is what is happening. The present moment feels special. It is real. However much you may remember the past or anticipate the future, you live in the present. Of course, the moment during which you read that sentence is no longer happening. This one is. In other words, it feels as though time flows, in the sense that the present is constantly updating itself. We have a deep intuition that the future is open until it becomes present and that the past is fixed. As time flows, this structure of fixed past, immediate present and open future gets carried forward in time. This structure is built into our language, thought and behavior. How we live our lives hangs on it.

Yet as natural as this way of thinking is, you will not find it reflected in science. The equations of physics do not tell us which events are occurring right now—they are like a map without the "you are here" symbol. The present moment does not exist in them, and therefore neither does the flow of time. Additionally, Albert Einstein's theories of relativity suggest not only that there is no single special present but also that all moments are equally real [see "That Mysterious Flow," by Paul Davies, on page 8]. Fundamentally,

KEITH PETERS

© 2018 Scientific American

the future is no more open than the past.

The gap between the scientific understanding of time and our everyday understanding of time has troubled thinkers throughout history. It has widened as physicists have gradually stripped time of most of the attributes we commonly ascribe to it. Now the rift between the time of physics and the time of experience is reaching its logical conclusion, for many in theoretical physics have come to believe that time fundamentally does not even exist.

WITHOUT THINKING DEEPLY ABOUT TIME, PHYSICISTS MAY NEVER MAKE PROGRESS ON A UNIFIED THEORY.

The idea of a timeless reality is initially so startling that it is hard to see how it could be coherent. Everything we do, we do in time. The world is a series of events strung together by time. Anyone can see that my hair is graying, that objects move, and so on. We see change, and change is the variation of properties with respect to time. Without time, the world would be completely still. A timeless theory faces the challenge of explaining how we see change if the world is not really changing.

Recent research attempts to perform just this feat. Although time may not exist at a fundamental level, it may arise at higher levels—just as a table feels solid even though it is a swarm of particles composed mostly of empty space. Solidity is a collective, or emergent, property of the particles. Time, too, could be an emergent property of whatever the basic ingredients of the world are.

This concept of emergent time is potentially as revolutionary as the development of the theories of relativity and of quantum mechanics a century ago. Einstein said that the key step forward in developing relativity was his reconceptualization of time. As physicists pursue his dream of unifying relativity with quantum mechanics, they believe that time is again central. In 2008 the Foundational Questions Institute (FQXi) sponsored an essay contest on the nature of time, and a veritable who's who of modern physics weighed in. Many held that a unified theory will describe a timeless world. Others were loath to get rid of time. The one thing they agreed on was that without thinking deeply about time, progress on unification may well be impossible.

THE RISE AND FALL OF TIME

OUR RICH COMMONSENSICAL notions of time have suffered a withering series of demotions throughout the ages. Time has many jobs to do in physics, but as physics has progressed, these jobs have been outsourced one by one.

It may not be obvious at first, but Isaac Newton's laws of motion require time to have many specific features. All observers in principle agree on the sequence in which events happen. No matter when or where an event occurs, classical physics assumes that you can objectively say whether it happens before, after or simultaneously with any other event in the universe. Time therefore provides a complete ordering of all the events in the world. Simultaneity is absolute—an observer-independent fact. Furthermore, time must be continuous so that we can define velocity and acceleration.

Classical time must also have a notion of duration—what physicists call a metric—so that we can tell how far apart in time events are from one another. To say that Olympic sprinter Usain Bolt can run as fast as 27 miles per hour, we need to have a measure of what an hour is. Like the order of events, duration is observer-independent. If Alice and Bob leave school at 3 P.M., go their separate ways, and then meet back at home at 6 P.M., the amount of time that has elapsed for Alice is equal to the amount of time that has elapsed for Bob.

In essence, Newton proposed that the world comes equipped with a master clock. The clock uniquely and objectively carves the world up into instants of time. Newton's physics listens to the ticking of this clock and no other. Newton additionally felt that time flows and that this flow gives us an arrow telling us which direction is the future, although these extra features are not strictly demanded by his laws.

Newton's time may seem old hat to us now, but a moment's reflection reveals how astonishing it is. Its many features—order, continuity, duration, simultaneity, flow and the arrow—are logically detachable, yet they all stick together in the master clock that Newton dubbed "time." This assembly of features succeeded so well that it survived unscathed for almost two centuries.

Then came the assaults of the late 19th and early 20th centuries. The first was the work of Austrian physicist Ludwig Boltzmann, who reasoned that, because Newton's laws work equally well going forward or backward in time, time has no built-in arrow. Instead he proposed that the distinction between past and future is not intrinsic to time but arises from asymmetries in how the matter in the universe is organized. Although physicists still debate the details of this proposal, Boltzmann convincingly plucked away one feature of Newtonian time.

Einstein mounted the next assault by doing away with the idea of absolute si-

IN BRIEF

Time is a perennial hot topic in physics. The search for a unified theory is forcing physicists to reexamine very basic assumptions, and few things are more basic than time.

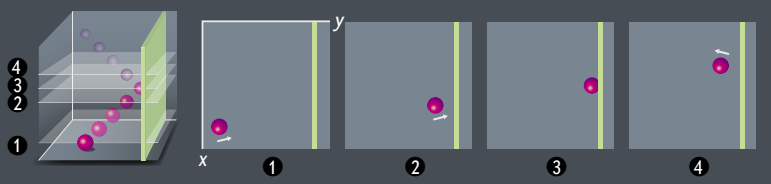
Some physicists argue that there is no such thing as time. Others think time ought to be promoted rather than demoted. In between these two positions is the fascinating idea that time exists but is not fundamental. A static world somehow gives rise to the time we perceive.

Philosophers have debated such ideas since before the time of Socrates, but physicists are now making them concrete. According to one, time may arise from the way that the universe is partitioned; what we perceive as time reflects the relations among its pieces.

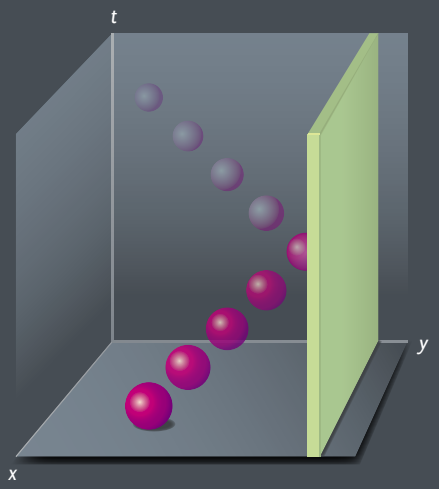
© 2018 Scientific American

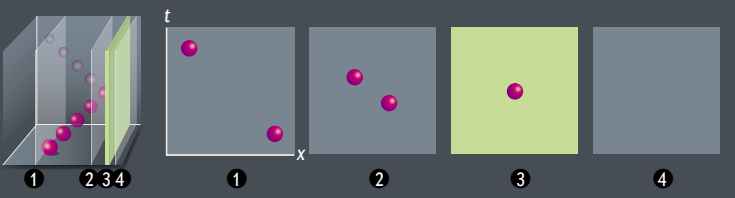
How Time Is Not Like Space

Physicists, artists and graph makers of all kinds routinely depict time as another dimension of space, creating a unified spacetime—shown here as a three-dimensional block in which a ball bounces off a wall. Relativity theory holds that spacetime can be sliced up in various ways. But not all are equally sensible.



The usual way takes slices of space at successive moments of time, creating a movie of the ball's motion. Each frame leads to the next, according to the familiar laws of physics.





An alternative considers slices not from past to future but from left to right. Each slice is part space, part time. To the left of the wall, the ball appears in two positions; on the right, it does not appear at all. If this slicing seems strange, it should: it makes the laws of physics very unwieldy.

multaneity. According to his special theory of relativity, what events are happening at the same time depends on how fast you are going. The true arena of events is not time or space, but their union: spacetime. Two observers moving at different velocities disagree on when and where an event occurs, but they agree on its spacetime location. Space and time are secondary concepts that, as mathematician Hermann Minkowski, who had been one of Einstein's university professors, famously declared, "are doomed to fade away into mere shadows."

Things only get worse in 1915 with Einstein's general theory of relativity, which extends special relativity to situations where the force of gravity operates. Gravity distorts time, so that a second's passage here may not mean the same thing as a second's passage there. Only in rare cases is it possible to synchronize clocks and have them stay synchronized, even in principle. You cannot generally think of the world as unfolding, tick by tick, according to a single time parameter. In extreme situations, the world might not be carvable into instants of time at all. It then becomes impossible to say that an event happened before or after another.

General relativity contains many functions with the English word "time" attached to them: coordinate time, proper time, global time. Together they perform many of the jobs Newton's single time did, but individually none of them seems worthy of the title. Either the physics does not listen to these clocks, or, if it does, those clocks apply only to small patches of the universe or to particular observers. Although physicists today fret that a unified theory will have to eliminate time, a good argument can be made that time was already lost by 1915 and that we just have not fully come to grips with it yet.

TIME AS THE GREAT STORYTELLER

WHAT GOOD IS TIME, THEN? You might be tempted to think that the difference between space and time has nearly vanished and that the true arena of events in a relativistic universe is a big four-dimensional block. Relativity appears to spatialize time: to turn it into merely one more direction within the block. Spacetime is like a loaf of bread that you can slice in different ways, called either "space" or "time" almost arbitrarily.

Yet even in general relativity, time retains a distinct and important function: namely, that of locally distinguishing between "timelike" and "spacelike" directions. Timelike-related events are those that can be causally related. An object or signal can pass from one event to the other, influencing what happens. Spacelike-related events are causally unrelated. No object or signal can get from one to the other. Mathematically, a mere minus sign differentiates the two directions, yet this minus sign has huge effects. Observers disagree on the sequence of spacelike events, but they all agree on the order of timelike events. If one observer perceives that an event can cause another, all observers do.

Illustrations by Melissa Thomas

© 2018 Scientific American

TIME MAY BE A CONVENIENT FICTION THAT NO MORE EXISTS IN THE NATURAL WORLD THAN MONEY DOES.

In my own essay for the FQXi contest in 2008, I explored what this feature of time means. Imagine slicing up spacetime from past to future; each slice is the 3-D totality of space at one instant of time. The sum of all these slices of spacelike-related events is 4-D spacetime. Alternatively, imagine looking at the world sideways and slicing it up accordingly. From this perspective, each 3-D slice is a strange amalgam of events that are spacelike-related (in just two dimensions) and timelike-related. These two methods of slicing are like carving up a loaf of bread either vertically or horizontally [*see box on preceding page*].

The first method is familiar to physicists, not to mention moviegoers. The frames of a movie represent slices of spacetime: they show space at successive moments of time. Like film aficionados who instantly figure out the plot and predict what happens next, physicists can take a single complete spatial slice and reconstruct what happens on the other spatial slices, simply by applying the laws of physics.

The second method of slicing has no simple analogy. It corresponds to carving up spacetime not from past to future but from east to west. An example of such a slice might be the north wall in your house plus what will happen on that wall in the future. From this slice, you apply the laws of physics to reconstruct what the rest of your house (and indeed the rest of the universe) looks like. If that sounds strange, it should. It is not immediately obvious whether the laws of physics let you do that. But as mathematician Walter Craig of McMaster University and philosopher Steven Weinstein of the University of Waterloo have shown, you can, at least in some simple situations.

Although both methods of slicing are possible in principle, they are profoundly different. In the normal, past-to-future slicing, the data you need to collect on a slice are fairly easy to obtain. For instance, you measure the velocities of all particles. The velocity of a particle in one location is independent of the velocity of a particle someplace else, making both of them straightforward to measure. But in the second method, the particles' properties are not independent; they have to be set up in a very specific way, or else a single slice would not suffice to reconstruct all the others. You would have to perform extremely difficult measurements on groups of particles to gather the data you need. To make matters worse, only in special cases, such as the one Craig and Weinstein discovered, would even these measurements allow you to reconstruct the full spacetime.

In a very precise sense, time is the direction within spacetime in which good prediction is possible—the direction in which we can tell the most informative stories. The narrative of the universe does not unfold in space. It unfolds in time.

QUANTUM TIME

ONE OF THE HIGHEST GOALS of modern physics is to unite general relativity with quantum mechanics, producing a single theory that handles both the gravitational and quantum aspects of matter—a quantum theory of gravity. One of the stumbling blocks has been that quantum mechanics requires time to have properties that contradict what I have said so far.

Quantum mechanics says that objects have a much richer repertoire of behaviors than we can possibly capture with classical quantities such as position and velocity. The full description of an object is given by a mathematical function called the quantum state. This state evolves continuously in time. Using it, physicists are able to calculate the probabilities of any experimental outcome at any time. If we send an electron through a device that will deflect it either up or down, quantum mechanics may not be able to tell us with certainty which outcome to expect. Instead the quantum state may give us only probabilities of outcomes; for instance, a 25 percent chance the electron will veer upward and a 75 percent chance it will veer downward. Two systems described with identical quantum states may give different outcomes. The outcomes of experiments are probabilistic.

The theory's probabilistic predictions require time to have certain features. First, time is that which makes contradictions possible. A rolled die cannot have both 5 and 3 facing up at the same time. It can do so only at different times. Connected to this feature is the fact that the probability of landing on each of the six numbers must add up to 100 percent; otherwise the concept of probability would not be meaningful. The probabilities add up at a time, not at a place. The same is true of the probabilities for quantum particles to have a given position or momentum.

Second, the temporal order of quantum measurements makes a difference. Suppose I pass an electron through a device that deflects it first along the vertical direction, then along the horizontal direction. As it emerges, I measure its angular momentum. I repeat the experiment, this time deflecting the electron horizontally, then vertically, and measuring its angular momentum again. The values I get will be vastly different.

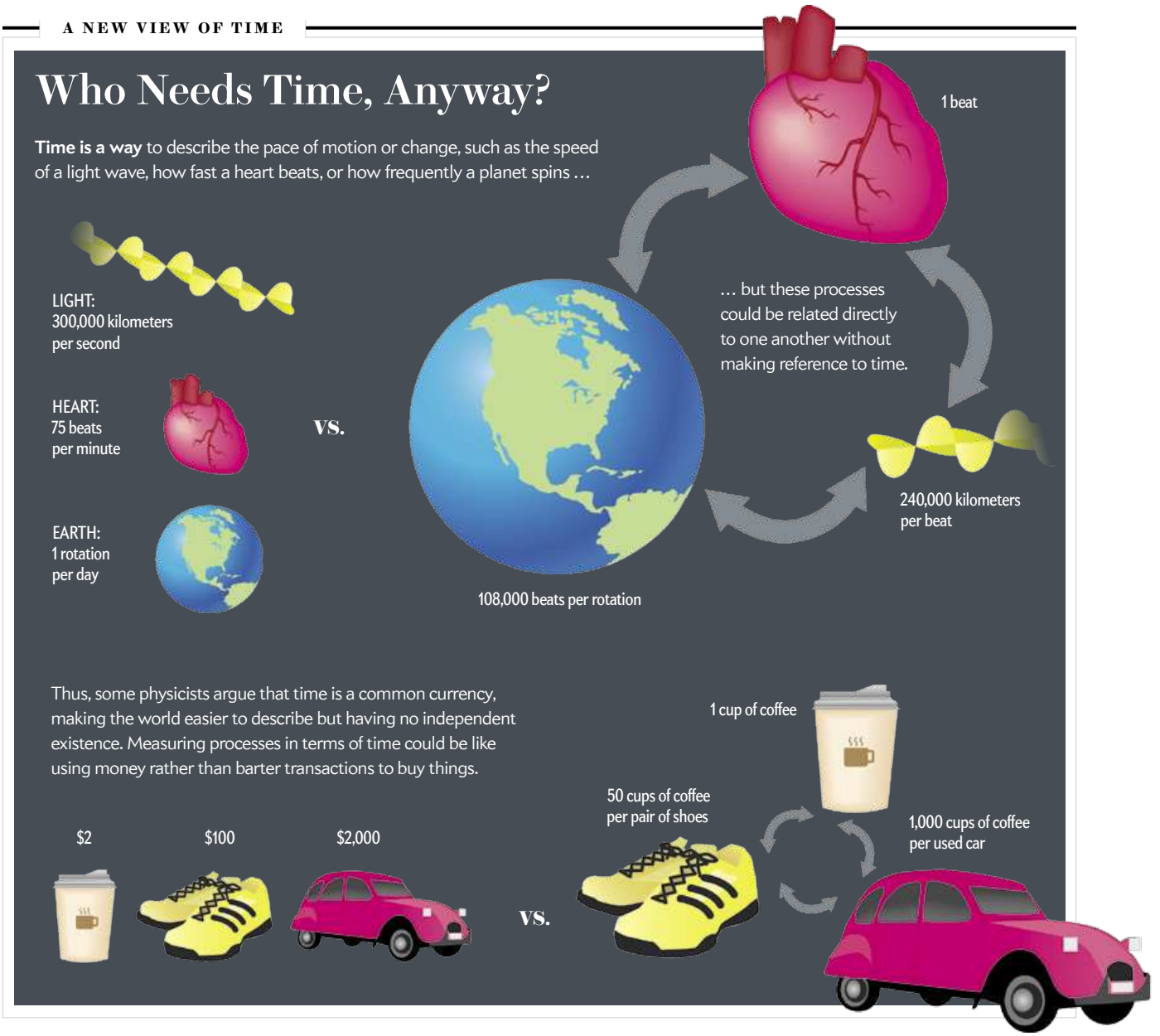
Third, a quantum state provides probabilities for all of space at an instant of time. If the state encompasses a pair of particles, then measuring one particle in-

CAUSAL SETS

Defenders of Time's Flow

Not all physicists think the world is fundamentally timeless. One of the more intriguing alternative ideas is causal set theory, developed by Rafael Sorkin and David Rideout of the Perimeter Institute for Theoretical Physics in Ontario. It supposes that the world is a set of events, called a causet, which grows as new events come into existence according to probabilistic rules. The hope is that the process reproduces the features of spacetime that we perceive, including the flow of time. An outstanding question, though, is whether this process produces worlds that are compatible with relativity theory. —*C.C.*

© 2018 Scientific American


Who Needs Time, Anyway?
Time is a way to describe the pace of motion or change, such as the speed of a light wave, how fast a heart beats, or how frequently a planet spins …
LIGHT: 300,000 kilometers per second
HEART: 75 beats per minute
VS.
EARTH: 1 rotation per day
1 beat
… but these processes could be related directly to one another without making reference to time.
240,000 kilometers per beat
108,000 beats per rotation
Thus, some physicists argue that time is a common currency, making the world easier to describe but having no independent existence. Measuring processes in terms of time could be like using money rather than barter transactions to buy things.
$2
$100
$2,000
VS.
1 cup of coffee
50 cups of coffee per pair of shoes
1,000 cups of coffee per used car

stantaneously affects the other no matter where it is—leading to the infamous "spooky action at a distance" that so troubled Einstein about quantum mechanics. The reason it bothered him was that for the particles to react at the same time, the universe must have a master clock, which relativity expressly forbids.

Although some of these issues are controversial, time in quantum mechanics is basically a throwback to time in Newtonian mechanics. Physicists fret about the absence of time in relativity, but perhaps a worse problem is the central role of time in quantum mechanics. It is the deep reason that unification has been so hard.

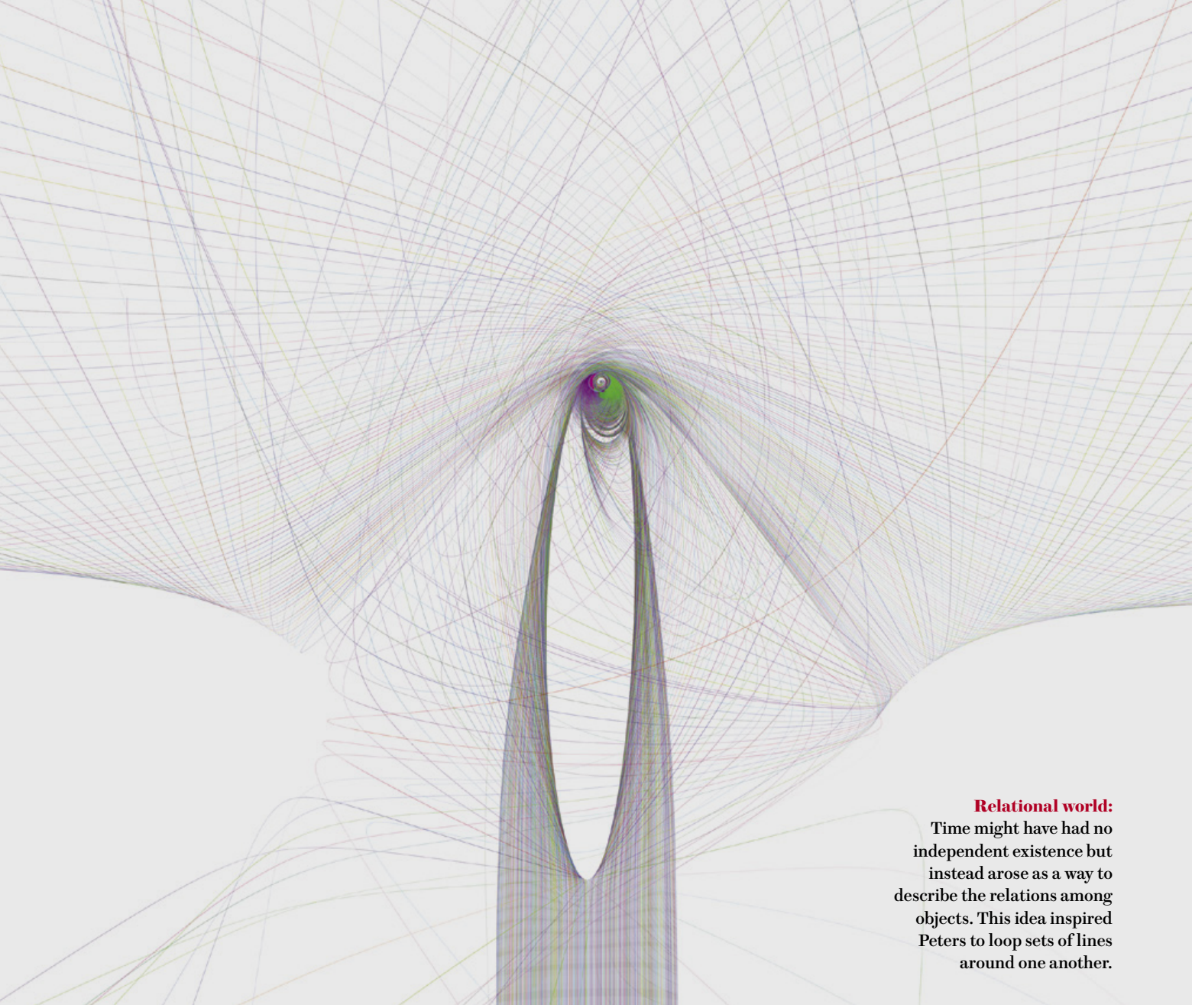
WHERE DID THE TIME GO?

A LARGE NUMBER of research programs have sought to reconcile general relativity and quantum mechanics: superstring theory, causal triangulation theory, noncommutative geometry, and more. They split roughly into two groups. Physicists who think quantum mechanics provides the firmer foundation, like superstring theorists, start with a full-blooded time. Those who believe that general relativity provides the better starting point begin with a theory in which time is already demoted and hence are more open to the idea of a timeless reality.

To be sure, the distinction between these two approaches is blurry. Superstring theorists have investigated timeless theories. To convey the basic problem that time poses, I will focus on the second approach. The leading instance of this strategy is loop quantum gravity [see "Atoms of Space and Time," by Lee Smolin, on page 94], which descends from an earlier program known as canonical quantum gravity.

Canonical quantum gravity emerged in the 1950s and 1960s, when physicists rewrote Einstein's equations for gravity in the same form as the equations for electromagnetism, the idea being that the same techniques used to develop a quan-

© 2018 Scientific American



Relational world: Time might have had no independent existence but instead arose as a way to describe the relations among objects. This idea inspired Peters to loop sets of lines around one another.

KEITH PETERS

tum theory of electromagnetism could then be applied to gravity as well. When the late physicists John Archibald Wheeler and Bryce DeWitt attempted this procedure in the late 1960s, they arrived at a very strange result. The equation (dubbed the Wheeler-DeWitt equation) utterly lacked a time variable. The symbol t denoting time had simply vanished.

Thus ensued decades of consternation among physicists. How could time just disappear? In retrospect, this result was not entirely surprising. As I mentioned earlier, time had already nearly disappeared from general relativity even before physicists attempted to merge it with quantum mechanics.

If you take this result literally, time does not really exist. Carlo Rovelli of Aix-Marseille University in France, one of the founders of loop quantum gravity, entitled his FQXi essay "Forget Time." He and English physicist Julian Barbour are the most prominent proponents of this idea. They have attempted to rewrite quantum mechanics in a timeless manner, as relativity appears to require.

The reason they think this maneuver is possible is that although general relativity lacks a global time, it still manages to describe change. In essence, it does so by relating physical systems directly to one another rather than to some abstract notion of global time. In Einstein's thought experiments, observers establish the timing of events by comparing clocks using light signals. We might describe the variation in the location of a satellite around Earth in terms of the ticks of a clock in my kitchen, or vice versa. What we are doing is describing the correlations between two physical objects, minus any global time as intermediary. Instead of describing my hair color as changing with time, we can correlate it with the satellite's orbit. Instead of saying a baseball accelerates at 10 meters per second per second, we can describe it in terms of the change of a glacier. And so on. Time becomes redundant. Change can be described without it.

This vast network of correlations is

© 2018 Scientific American

neatly organized so that we can define something called "time" and relate everything to it, relieving ourselves of the burden of keeping track of all those direct relations. Physicists are able to compactly summarize the workings of the universe in terms of physical laws that play out in time. Yet this convenient fact should not trick us into thinking that time is a fundamental part of the world's furniture. Money, too, makes life much easier than negotiating a barter transaction every time you want to buy coffee, although it is an invented placeholder for the things we value, not something we value in and of itself. Similarly, time allows us to relate physical systems to one another without trying to figure out exactly how a glacier relates to a baseball. But it, too, may be a convenient fiction that no more exists fundamentally in the natural world than money does.

Getting rid of time has its appeal but inflicts a good deal of collateral damage. For one, it requires quantum mechanics to be thoroughly rethought. Consider the famous case of Schrödinger's cat. The cat is suspended between life and death, its fate hinging on the state of a quantum particle. In the usual way of thinking, the cat becomes one or the other after a measurement or some equivalent process takes place. Rovelli, though, would argue that the status of the cat is never resolved. The poor thing may be dead with respect to itself, alive relative to a human in the room, dead relative to a second human outside the room, and so on.

It is one thing to make the timing of the cat's death depend on the observer, as special relativity does. It is rather more surprising to make whether it even happens relative, as Rovelli suggests, following the spirit of relativity as far as it will go. Because time is so basic, banishing it would transform physicists' worldview.

THE RECOVERY OF TIME

EVEN IF THE WORLD is fundamentally timeless, still it seems as though it does have time in it. An urgent question for anyone espousing timeless quantum gravity is explaining why the world seems temporal. General relativity, too, lacks Newtonian time, but at least it has various partial substitutes that together behave like Newtonian time when gravity is weak and relative velocities low. The Wheeler-DeWitt equation lacks even those substitutes. Barbour and Rovelli have each offered suggestions for how time (or at least the illusion of time) could pop out of nothingness. But canonical quantum gravity already offers a more developed idea.

Known as semiclassical time, it goes back to a 1931 paper by English physicist Nevill F. Mott that described the collision between a helium nucleus and a larger atom. To model the total system, Mott applied an equation that lacks time and usually is applied only to static systems. He then divided the system into two subsystems and used the helium nucleus as a "clock" for the atom. Remarkably, the atom, relative to the nucleus, obeys the standard time-dependent equation of quantum mechanics. A function of space plays the role of time. So even though the system as a whole is timeless, the individual pieces are not. Hidden in the timeless equation for the total system is a time for the subsystem.

Much the same works for quantum gravity, as Claus Kiefer of the University of Cologne in Germany, following in the footsteps of Thomas Banks of the University of California, Santa Cruz, and others, argued in his FQXi essay. The universe may be timeless, but if you imagine breaking it into pieces, some of the pieces can serve as clocks for the others. Time emerges from timelessness. We perceive time because we are, by our very nature, one of those pieces.

As interesting and startling as this idea is, it leaves us wanting more. The universe cannot always be broken up into pieces that serve as clocks, and in those cases, the theory makes no probabilistic predictions. Handling those situations will take a full quantum theory of gravity and a deeper rethinking of time.

Historically, physicists began with the highly structured time of experience, the time of a fixed past, a present and an open future. They gradually dismantled this structure, and little, if any, of it remains. Researchers must now reverse this train of thought and reconstruct the time of experience from the time of nonfundamental physics, which itself may need to be reconstructed from a network of correlations among pieces of a fundamental static world.

French philosopher Maurice Merleau-Ponty argued that time itself does not really flow and that its apparent flow is a product of our "surreptitiously putting into the river a witness of its course." That is, the tendency to believe time flows is a result of forgetting to put ourselves and our connections to the world into the picture. Merleau-Ponty was speaking of our subjective experience of time, and until recently no one ever guessed that objective time might itself be explained as a result of those connections. Time may exist only by breaking the world into subsystems and looking at what ties them together. In this picture, physical time emerges by virtue of our thinking ourselves as separate from everything else. SA

TIME MAY EXIST ONLY BY BREAKING THE WORLD INTO SUBSYSTEMS AND LOOKING AT WHAT TIES THEM TOGETHER.

MORE TO EXPLORE

Relational Quantum Mechanics. Carlo Rovelli in *International Journal of Theoretical Physics*, Vol. 35, No. 8, pages 1637–1678; August 1996. http://arxiv.org/abs/quant-ph/9609002

Can Physics Coherently Deny the Reality of Time? Richard Healey in *Time, Reality & Experience*. Edited by Craig Callender. Cambridge University Press, 2002.

Three Roads to Quantum Gravity. Lee Smolin. Basic Books, 2002.

The Fabric of the Cosmos: Space, Time and the Texture of Reality. Brian Greene. Vintage, 2005.

The Complete Idiot's Guide to String Theory. George Musser. Alpha, 2008.

From Eternity to Here: The Quest for the Ultimate Theory of Time. Sean M. Carroll. Dutton, 2010.

Does Time Exist in Quantum Gravity? Claus Kiefer. www.fqxi.org/community/forum/topic/265

Forget Time. Carlo Rovelli. www.fqxi.org/community/forum/topic/237

The Nature of Time. Julian Barbour. www.fqxi.org/community/forum/topic/360

What Makes Time Special. Craig Callender. www.fqxi.org/community/forum/topic/302

scientificamerican.com/magazine/sa

© 2018 Scientific American

George Musser is a contributing editor for *Scientific American* and author of *The Complete Idiot's Guide to String Theory* (Alpha, 2008) and of *Spooky Action at a Distance* (Farrar, Straus and Giroux, 2015).

PHILOSOPHY

A HOLE AT THE HEART OF PHYSICS

Physicists can't seem to find the time—literally. Can philosophers help?

By George Musser

FOR MOST PEOPLE, THE GREAT MYSTERY OF TIME IS THAT there never seems to be enough of it. If it is any consolation, physicists are having much the same problem. The laws of physics contain a time variable, but it fails to capture key aspects of time as we live it—notably, the distinction between past and future. And as researchers try to formulate more fundamental laws, the little *t* evaporates altogether. Stymied, many physicists have sought help from an unfamiliar source: philosophers.

IN BRIEF

Physicists have turned to philosophers for help with merging quantum mechanics and relativity, which treat space and time in two very different ways, into a single theory of quantum gravity.

By revisiting a thought experiment created by Albert Einstein, philosophers have found that general relativity has puzzling implications for the question of whether space and time are entities in their own right or merely artificial devices to describe the relations between physical objects.

Philosophers are also helping to identify inconsistencies in physicists' theories by, for example, pointing out that attempts to explain the arrow of time—the asymmetry of past and future—often rely on circular reasoning.

From philosophers? To most physicists, that sounds rather quaint. The closest some get to philosophy is a late-night conversation over dark beer. Even those who have read serious philosophy generally doubt its usefulness; after a dozen pages of Immanuel Kant, philosophy begins to seem like the unintelligible in pursuit of the undeterminable. "To tell you the truth, I think most of my colleagues are terrified of talking to philosophers—like being caught coming out of a pornographic cinema," says physicist Max Tegmark of the Massachusetts Institute of Technology.

It was not always so. Philosophers played a crucial role in past scientific revolutions, including the development of quantum mechanics and relativity in the early 20th century. Today a new revolution is under way, as physicists struggle to merge those two theories into a theory of quantum gravity—a theory that will have to reconcile two vastly different conceptions of space and time. Carlo Rovelli of Aix-Marseille University in France, a leader in this effort, observes that "the contributions of philosophers to the new understanding of space and time in quantum gravity will be very important."

Two examples illustrate how physicists and philosophers have been pooling their resources. The first concerns the "problem of frozen time," also known simply as the "problem of time." It arises when theorists try to turn Einstein's general theory of relativity into a quantum theory using a procedure called canonical quantization. The procedure worked brilliantly when applied to the theory of electromag-

© 2018 Scientific American

netism, but in the case of relativity, it produces an equation—the Wheeler-DeWitt equation—without a time variable. Taken literally, the equation predicts that the universe should be frozen in time, in direct contradiction with what we see.

DON'T LOSE ANY MORE TIME

THIS UNHAPPY OUTCOME may reflect a flaw in the procedure itself, but some physicists and philosophers argue that it has deeper roots, right down to one of the founding principles of relativity: that the laws of physics are the same for all observers. Physicists think of the principle in geometric terms. Two observers will perceive spacetime to have two different shapes, corresponding to their views of who is moving and what forces are acting. Each shape is a smoothly warped version of the other, in the way that a coffee cup is a reshaped doughnut. Relativity theory says that the difference cannot be meaningful. Therefore, any two such shapes are physically equivalent.

In the late 1980s philosophers John Earman and John D. Norton, both at the University of Pittsburgh, argued that general relativity has startling implications for an old metaphysical question: Do space and time exist independently of stars, galaxies and their other contents (a position known as substantivalism), or are they merely an artificial device to describe how physical objects are related (relationism)? As Norton has written: "Are they like a canvas onto which an artist paints; they exist whether or not the artist paints on them? Or are they akin to parenthood; there is no parenthood until there are parents and children."

He and Earman revisited a long-neglected thought experiment of Albert Einstein's. Consider an empty patch of spacetime. Outside this hole the distribution of matter fixes the geometry of spacetime, per the equations of relativity. Inside, however, the equations let spacetime take on any of a variety of shapes. Spacetime behaves like a canvas tent. The tent poles, which represent matter, force the canvas to assume a certain shape. If you leave out a pole, however, creating the equivalent of a hole, part of the tent can sag, or bow out, or ripple unpredictably in the wind. A similar ambiguity arises everywhere, not just in holes.

The thought experiment poses a dilemma. If the continuum is a thing in its own right (as substantivalism holds), general relativity must be indeterministic—that is, its description of the world must contain an element of randomness. For the theory to be deterministic, spacetime must be a mere fiction (as relationism holds). At first glance, it looks like a victory for relationism. It helps that other theories, such as electromagnetism, are based on symmetries that resemble relationism.

DO SPACE AND TIME EXIST INDEPENDENTLY OF STARS, GALAXIES AND OTHER OBJECTS, OR ARE THEY MERELY AN ARTIFICIAL DEVICE?

But relationism has its own troubles. It is the ultimate source of the problem of frozen time: space may morph over time, but if its many shapes are all equivalent, it never truly changes. Moreover, relationism clashes with the substantivalist underpinnings of quantum mechanics. If spacetime has no fixed meaning, how can you make observations at specific places and moments, as quantum mechanics seems to require?

Different resolutions of the dilemma lead to very different theories of quantum gravity. Some, such as Rovelli and English physicist Julian Barbour, are trying a relationist approach; they think time does not exist and have searched for ways to explain change as an illusion. Others, including string theorists, lean toward substantivalism.

"It's a good example of the value of philosophy of physics," says philosopher Craig Callender of the University of California, San Diego. "If physicists think the problem of time in canonical quantum gravity is solely a quantum problem, they're hurting their understanding of the problem—for it's been with us for much longer and is more general."

RUNNING ON ENTROPY

A SECOND EXAMPLE of philosophers' contributions concerns the arrow of time—the asymmetry of past and future. Many people assume that the arrow is explained by the second law of thermodynamics, which states that entropy, loosely defined as the amount of disorder within a system, increases with time. Yet no one can really account for the second law.

The leading explanation, put forward by 19th-century Austrian physicist Ludwig Boltzmann, is probabilistic. The basic idea is that there are more ways for a system to be disordered than to be ordered. If the system is fairly ordered now, it will probably be more disordered a moment from now. This reasoning, however, is symmetric in time. The system was probably more disordered a moment ago, too. As Boltzmann recognized, the only way to ensure that entropy will increase into the future is if it starts off with a low value in the past. Thus, the second law is not so much a fundamental truth as historical happenstance, perhaps related to events early in the big bang.

Philosopher Huw Price of the University of Cambridge argues that almost every attempt to explain time asymmetry suffers from circular reasoning—some hidden presumption of time asymmetry. This fallacy also distorts how physicists interpret quantum mechanics. Price's work is an example of how philosophers can serve, in the words of philosopher Richard Healey of the University of Arizona, as the "intellectual conscience of the practicing physicist." Trained in logical rigor, they are experts at tracking down subtle biases.

Life would be boring if we always listened to our conscience, and physicists have often done best when ignoring philosophers. Yet in the eternal battle against our own leaps of logic, conscience is sometimes all we have to go on. SA

MORE TO EXPLORE

Time's Arrow and Archimedes' Point: New Directions for the Physics of Time. Huw Price. Oxford University Press, 1997.

Quantum Spacetime: What Do We Know? Carlo Rovelli in *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity.* Edited by Craig Callender and Nick Huggett. Cambridge University Press, 2001. http://arxiv.org/abs/gr-qc/9903045

From Metaphysics to Physics. Gordon Belot and John Earman in *From Physics to Philosophy.* Edited by Jeremy Butterfield and Constantine Pagonis. Reprint edition. Cambridge University Press, 2010.

scientificamerican.com/magazine/sa

© 2018 Scientific American

Paul Davies is a theoretical physicist and cosmologist at Arizona State University, where he is director of BEYOND: Center for Fundamental Concepts in Science. He is one of the most prolific writers of popular-level books in physics.

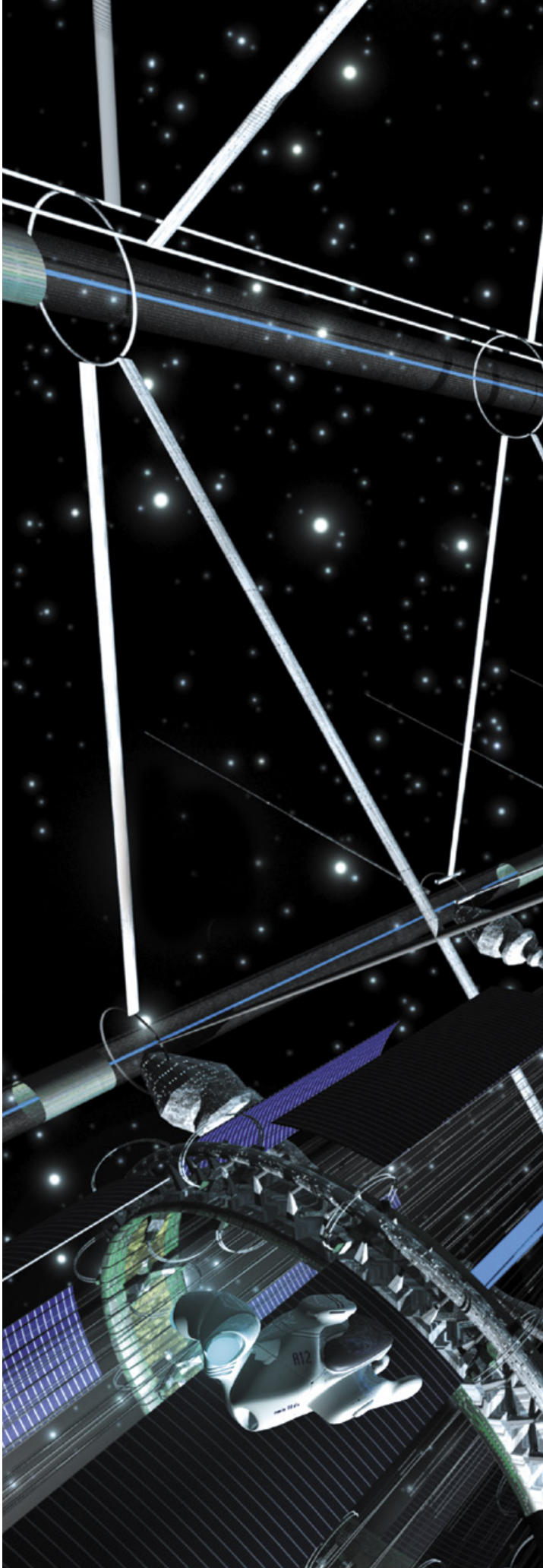
TRAVEL

How to Build a Time Machine

It wouldn't be easy, but it might be possible

By Paul Davies

Illustration by Peter Bollinger



Wormhole generator/towing machine is imagined by futurist artist Peter Bollinger. This painting depicts a gigantic space-based particle accelerator that is capable of creating, enlarging and moving wormholes for use as time machines.

© 2018 Scientific American

IN BRIEF

Traveling forward in time is easy enough. If you move close to the speed of light or sit in a strong gravitational field, you experience time more slowly than other people do—another way of saying that you travel into their future.

Traveling into the past is rather trickier. Relativity theory allows it in certain spacetime configurations: a rotating universe, a rotating cylinder and, most famously, a wormhole—a tunnel through space and time.

© 2018 Scientific American

TIME TRAVEL HAS BEEN A POPULAR SCIENCE-FICTION THEME SINCE H. G. WELLS wrote his celebrated novel *The Time Machine* in 1895. But can it really be done? Is it possible to build a machine that would transport a human being into the past or future?

For decades time travel lay beyond the fringe of respectable science. In recent years, however, the topic has become something of a cottage industry among theoretical physicists. The motivation has been partly recreational—time travel is fun to think about. But this research has a serious side, too. Understanding the relation between cause and effect is a key part of attempts to construct a unified theory of physics. If unrestricted time travel were possible, even in principle, the nature of such a unified theory could be drastically affected.

Our best understanding of time comes from Albert Einstein's theories of relativity. Prior to these theories, time was widely regarded as absolute and universal, the same for everyone no matter what their physical circumstances were. In his special theory of relativity, Einstein proposed that the measured interval between two events depends on how the observer is moving. Crucially, two observers who move differently will experience different durations between the same two events.

The effect is often described using the "twin paradox." Suppose that Sally and Sam are twins. Sally boards a rocket ship and travels at high speed to a nearby star, turns around and flies back to Earth, while Sam stays at home. For Sally the duration of the journey might be, say, one year, but when she returns and steps out of the spaceship, she finds that 10 years have elapsed on Earth. Her brother is now nine years older than she is. Sally and Sam are no longer the same age, despite the fact that they were born on the same day. This example illustrates a limited type of time travel. In effect, Sally has leaped nine years into Earth's future.

EXISTING FORMS OF FORWARD TIME TRAVEL

SYSTEM	SPECIFICATIONS	CUMULATIVE TIME LAG
Airline flight	920 kilometers per hour for eight hours	10 nanoseconds (relative to inertial reference frame)
Nuclear submarine tour	300 meters' depth for six months	500 nanoseconds (relative to sea level)
Cosmic-ray neutron	10^{18} electron volts	Mean life stretched from 15 minutes to 30,000 years
Neutron star	Redshift 0.3	Time intervals expand 30 percent (relative to deep space)

JET LAG

THE EFFECT, KNOWN AS TIME DILATION, occurs whenever two observers move relative to each other. In daily life we don't notice weird time warps, because the effect becomes dramatic only when the motion occurs at close to the speed of light. Even at aircraft speeds, the time dilation in a typical journey amounts to just a few nanoseconds—hardly an adventure of Wellsian proportions. Nevertheless, atomic clocks are accurate enough to record the shift and confirm that time really is stretched by motion. So travel into the future is a proved fact, even if it has so far been in rather unexciting amounts.

To observe really dramatic time warps, one has to look beyond the realm of ordinary experience. Subatomic particles can be propelled at nearly the speed of light in large accelerator machines. Some of these particles, such as muons, have a built-in clock because they decay with a definite half-life; in accordance with Einstein's theory, fast-moving muons inside accelerators are observed to decay in slow motion. Some cosmic rays also experience spectacular time warps. These particles move so close to the speed of light that, from their point of view, they cross the galaxy in minutes, even though in Earth's frame of reference they seem to take tens of thousands of years. If time dilation did not occur, those particles would never make it here.

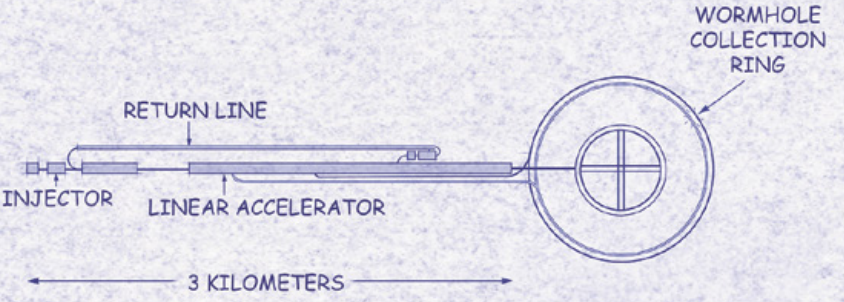
Speed is one way to jump ahead in time. Gravity is another. In his general theory of relativity, Einstein predicted that gravity slows time. Clocks run a bit faster in the attic than in the basement, which is closer to the center of Earth and therefore deeper down in a gravitational field. Similarly, clocks run faster in space than on the ground. Once again the effect is minuscule, but it has been directly measured using accurate clocks. Indeed, these time-warping effects have to be taken into account in the Global Positioning System. If they weren't, sailors, taxi drivers and cruise missiles could find themselves many kilometers off course.

At the surface of a neutron star, gravity is so strong that time is slowed by about 30 percent relative to Earth time. Viewed from such a star, events here would resemble a fast-forwarded video. A black hole represents the ultimate time warp; at the surface of the hole, time stands

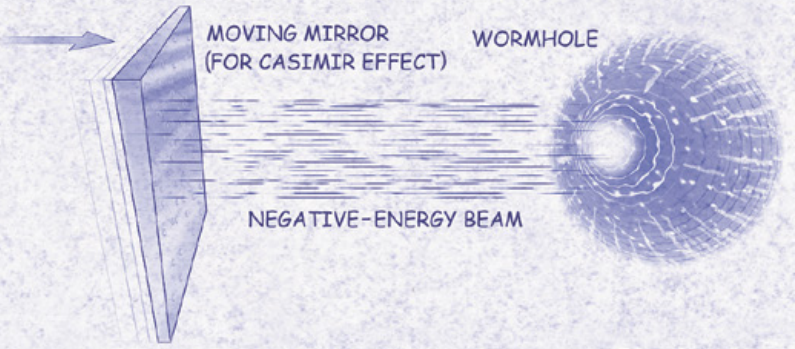
© 2018 Scientific American

A Wormhole Time Machine in Three Not So Easy Steps

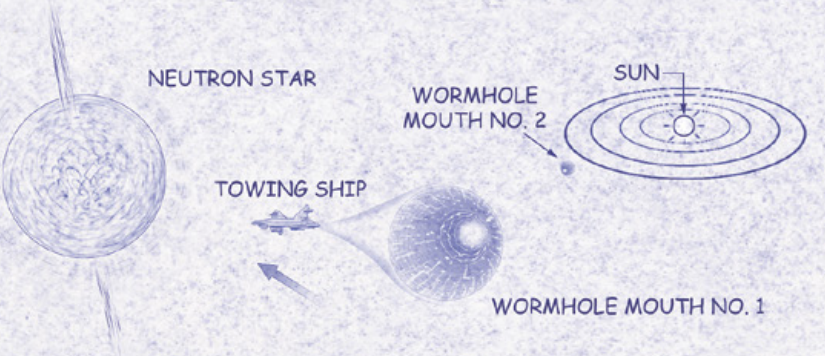
1 Find or build a wormhole—a tunnel connecting two different locations in space. Large wormholes might exist naturally in deep space, a relic of the big bang. Otherwise we would have to make do with subatomic wormholes—either natural ones (which are thought to be winking in and out of existence all around us) or artificial ones (produced by particle accelerators, as imagined here). These smaller wormholes would have to be enlarged to useful size, perhaps using energy fields like those that caused space to inflate shortly after the big bang.



2 Stabilize the wormhole. An infusion of negative energy, produced by quantum means such as the so-called Casimir effect, would allow a signal or object to pass safely through the wormhole. Negative energy counteracts the tendency of the wormhole to pinch off into a point of infinite or near-infinite density. In other words, it prevents the wormhole from becoming a black hole.



3 Tow the wormhole. A spaceship, presumably of highly advanced technology, would separate the mouths of the wormhole. One mouth might be positioned near the surface of a neutron star, an extremely dense star with a strong gravitational field. The intense gravity causes time to pass more slowly. Because time passes more quickly at the other wormhole mouth, the two mouths become separated not only in space but also in time.



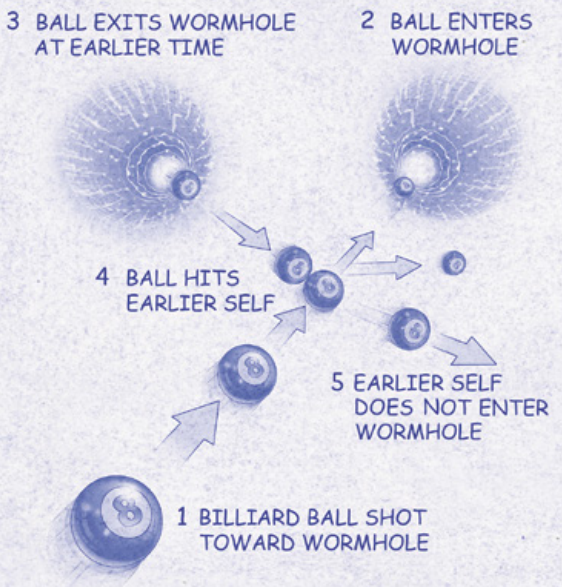
Illustrations by Philip Howe

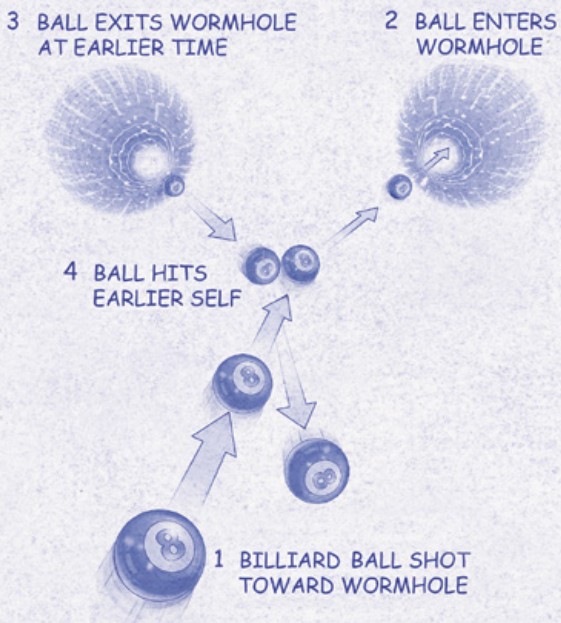
© 2018 Scientific American

Mother of All Paradoxes

The notorious mother paradox (sometimes formulated using other familial relationships) arises when people or objects can travel backward in time and alter the past. A simplified version involves billiard balls. A billiard ball passes through a wormhole time machine. Upon emerging, it hits its earlier self, thereby preventing it from ever entering the wormhole.

Resolution of the paradox proceeds from a simple realization: the billiard ball cannot do something that is inconsistent with logic or with the laws of physics. It cannot pass through the wormhole in such a way that will prevent it from passing through the wormhole. Yet nothing stops it from passing through the wormhole in an infinity of other ways.


3 BALL EXITS WORMHOLE AT EARLIER TIME
2 BALL ENTERS WORMHOLE
4 BALL HITS EARLIER SELF
5 EARLIER SELF DOES NOT ENTER WORMHOLE
1 BILLIARD BALL SHOT TOWARD WORMHOLE


3 BALL EXITS WORMHOLE AT EARLIER TIME
2 BALL ENTERS WORMHOLE
4 BALL HITS EARLIER SELF
1 BILLIARD BALL SHOT TOWARD WORMHOLE

still relative to Earth. This means that if you fell into a black hole from nearby, in the brief interval it took you to reach the surface, all of eternity would pass by in the wider universe. The region within the black hole is therefore beyond the end of time, as far as the outside universe is concerned. If an astronaut could zoom very close to a black hole and return unscathed—admittedly a fanciful, not to mention foolhardy, prospect—he could leap far into the future.

MY HEAD IS SPINNING

THUS FAR I HAVE DISCUSSED traveling forward in time. What about going backward? This is much more problematic. In 1948 Kurt Gödel of the Institute for Advanced Study in Princeton, N.J., produced a solution of Einstein's gravitational field equations that described a rotating universe. In this universe, an astronaut could travel through space so as to reach his own past. This comes about because of the way gravity affects light. The rotation of the universe would drag light (and thus the causal relations between objects) around with it, enabling a material object to travel in a closed loop in space that is also a closed loop in time, without at any stage exceeding the speed of light in the immediate neighborhood of the particle. Gödel's solution was shrugged aside as a mathematical curiosity—after all, observations show no sign that the universe as a whole is spinning. His result nonetheless served to demonstrate that going back in time was not forbidden by the theory of relativity. Indeed, Einstein confessed that he was troubled by the thought that his theory might permit travel into the past under some circumstances.

Other scenarios have been found to permit travel into the past. For example, in 1974 Frank J. Tipler of Tulane University calculated that a massive, infinitely long cylinder spinning on its axis at near the speed of light could let astronauts visit their own past, again by dragging light around the cylinder into a loop. In 1991 J. Richard Gott of Princeton University predicted that cosmic strings—structures that cosmologists think were created in the early stages of the big bang—could produce similar results. Then, in the mid-1980s, the most realistic scenario for a time machine emerged, based on the concept of a wormhole.

In science fiction, wormholes are sometimes called stargates; they offer a shortcut between two widely separated points in space. Jump through a hypothetical wormhole, and you might come out moments later on the other side of the galaxy. Wormholes naturally fit into the general theory of relativity, whereby gravity warps not only time but also space. The theory allows the analogue of alternative road and tunnel routes connecting

© 2018 Scientific American

two points in space. Mathematicians refer to such a space as multiply connected. Just as a tunnel passing under a hill can be shorter than the surface street, a wormhole may be shorter than the usual route through ordinary space.

The wormhole was used as a fictional device by Carl Sagan in his 1985 novel *Contact*. Prompted by Sagan, Kip S. Thorne and his co-workers at the California Institute of Technology set out to find whether wormholes were consistent with known physics. Their starting point was that a wormhole would resemble a black hole in being an object with fearsome gravity. Yet unlike a black hole, which offers a one-way journey to nowhere, a wormhole would have an exit as well as an entrance.

IN THE LOOP

FOR THE WORMHOLE to be traversable, it must contain what Thorne termed exotic matter. In effect, this is something that will generate antigravity to combat the natural tendency of a massive system to implode into a black hole under its intense weight. Antigravity, or gravitational repulsion, can be generated by negative energy or pressure. Negative-energy states are known to exist in certain quantum systems, which suggests that Thorne's exotic matter is not ruled out by the laws of physics, although it is unclear whether enough antigravitating stuff can be assembled to stabilize a wormhole.

Soon Thorne and his colleagues realized that if a stable wormhole could be created, then it could readily be turned into a time machine. An astronaut who passed through one might come out not only somewhere else in the universe but somewhen else, too—in either the future or the past.

To adapt the wormhole for time travel, one of its mouths could be towed to a neutron star and placed close to its surface. The gravity of the star would slow time near that wormhole mouth, so a time difference between the ends of the wormhole would gradually accumulate. If both mouths were then parked at a convenient place in space, this time difference would remain frozen in.

Suppose the difference were 10 years. An astronaut passing through the wormhole in one direction would jump 10 years into the future, whereas an astronaut passing in the other direction would jump 10 years into the past. By returning to his starting point at high speed across ordinary space, the second astronaut might get back home before he left. In other words, a closed loop in space could become a loop in time as well. The one restriction is that the astronaut could not return to a time before the wormhole was first built.

A formidable problem that stands in the way of making a wormhole time machine is the creation of the wormhole in the first place. Possibly space is threaded with such structures naturally—relics of the big bang. If so, a supercivilization might commandeer one. Alternatively, wormholes might naturally come into existence on tiny scales, the so-called Planck length, about 20 factors of 10 as small as an atomic nucleus. In principle, such a minute wormhole could be stabilized by a pulse of energy and then somehow inflated to usable dimensions.

CENSORED!

ASSUMING THAT THE ENGINEERING problems could be overcome, the production of a time machine could open up a Pandora's box of causal paradoxes. Consider, for example, the time traveler who visits the past and murders his mother when she was a young girl. How do we make sense of this? If the girl dies, she cannot become the time traveler's mother. If the time traveler was never born, however, he could not go back and murder his mother.

Paradoxes of this kind arise when the time traveler tries to change the past, which is obviously impossible. But that does not prevent someone from being a part of the past. Suppose the time traveler goes back and rescues a young girl from murder, and this girl grows up to become his mother. The causal loop is now self-consistent and no longer paradoxical. Causal consistency might impose restrictions on what a time traveler is able to do, but it does not rule out time travel per se.

Even if time travel is not strictly paradoxical, it is certainly weird. Consider the time traveler who leaps ahead a year and reads about a new mathematical theorem in a future edition of *Scientific American*. He notes the details, returns to his own time and teaches the theorem to a student, who then writes it up for *Scientific American*. The article is, of course, the very one that the time traveler read. The question then arises: Where did the information about the theorem come from? Not from the time traveler, because he read it, but not from the student either, who learned it from the time traveler. The information seemingly came into existence from nowhere, reasonlessly.

The bizarre consequences of time travel have led some scientists to reject the notion outright. Stephen Hawking of the University of Cambridge has proposed a "chronology protection conjecture," which would outlaw causal loops. Because the theory of relativity is known to permit causal loops, chronology protection would require some other factor to intercede to prevent travel into the past. What might this factor be? One suggestion is that quantum processes will come to the rescue. The existence of a time machine would allow particles to loop into their own past. Calculations hint that the ensuing disturbance would become self-reinforcing, creating a runaway surge of energy that would wreck the wormhole.

Chronology protection is still just a conjecture, so time travel remains a possibility. A final resolution of the matter may have to await the successful union of quantum mechanics and gravitation, perhaps through a theory such as string theory or its extension, so-called M-theory. It is even conceivable that the next generation of particle accelerators—or perhaps the existing Large Hadron Collider at CERN near Geneva—will be able to produce subatomic collisions powerful enough to rip spacetime itself, creating wormholes that survive long enough for nearby particles to execute fleeting causal loops. This would be a far cry from Wells's vision of a time machine, but it would forever change our picture of physical reality. SA

MORE TO EXPLORE

Time Machines: Time Travel in Physics, Metaphysics, and Science Fiction. Paul J. Nahin. American Institute of Physics, 1993.

The Quantum Physics of Time Travel. David Deutsch and Michael Lockwood in *Scientific American*, Vol. 270, No. 3, pages 68–74; March 1994.

Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1995.

Time Travel in Einstein's Universe: The Physical Possibilities of Travel through Time. J. Richard Gott III. Mariner Books, 2002.

How to Build a Time Machine. Paul Davies. Penguin, 2003.

scientificamerican.com/magazine/sa

© 2018 Scientific American

Ronald C. Lasky is an instructional professor at Dartmouth College and a senior technologist at Indium Corporation. One of his hobbies is studying the subtleties of relativity and quantum theory so that he can explain them to laypeople.

THOUGHT EXPERIMENT

TIME AND THE TWIN PARADOX

Does time tick by at the same rate for everyone?

By Ronald C. Lasky

IN BRIEF

We take for granted that time ticks by at the same rate for everyone. But Einstein's special theory of relativity shows that this assumption is not strictly true.

The classic case of time disparity involves twins—one of whom leaves Earth and travels round-trip to a star at nearly the speed of light, arriving back much younger than his brother. This aging difference is noticeable only when long distances are traveled at speeds approximating the speed of light.

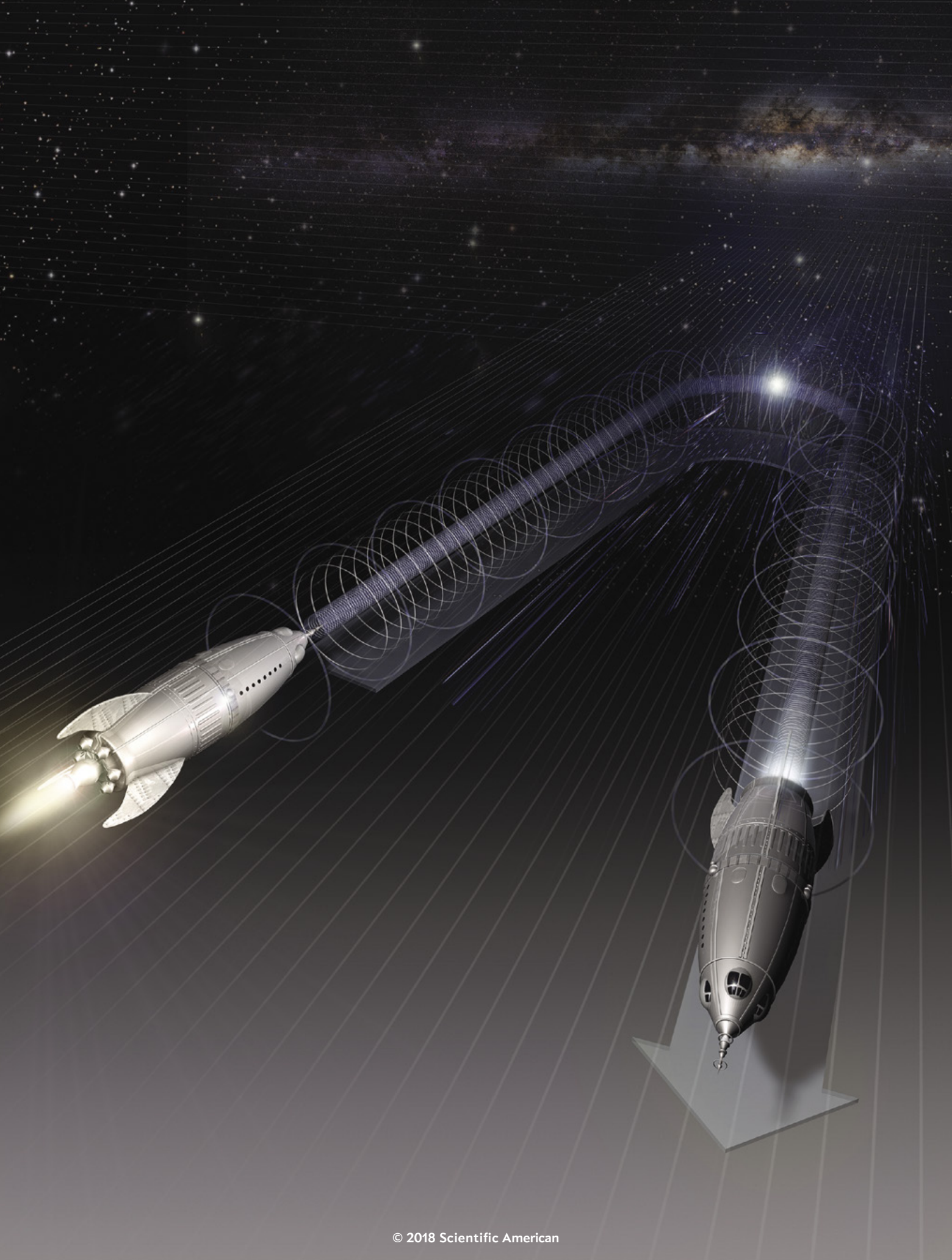
AS MAXIMS GO, "TIME IS RELATIVE" MAY NOT BE QUITE as famous as "time is money." But the notion that time speeds up or slows down depending on how fast one object is traveling relative to another surely ranks as one of Albert Einstein's most inspired insights.

The term "time dilation" was coined to describe the slowing of time caused by motion. To illustrate the effect of time dilation, Einstein proposed an example—the twin paradox—that is arguably the most famous thought experiment in relativity theory. In this supposed paradox, one of two twins travels at near the speed of light to a distant star and returns to Earth. Relativity dictates that when he comes back, he is younger than his identical twin [see "How to Build a Time Machine," by Paul Davies, on page 24].

The paradox lies in the question "Why is the traveling brother younger?" Special relativity tells us that an observed clock, travel-

Illustration by Malcolm Godwin

© 2018 Scientific American



© 2018 Scientific American

ing at high speed past an observer, appears to run more slowly—that is, it experiences time dilation. (Many of us solved this traveling-clock problem in sophomore physics to demonstrate one effect of the absolute nature of the speed of light.) Because special relativity says that there is no absolute motion, wouldn't the brother traveling to the star also see his brother's clock on Earth move more slowly? If this were the case, wouldn't they both be the same age?

This paradox is discussed in many books but solved in very few. It is typically explained by saying that the one who feels the acceleration is the one who is younger at the end of the trip; hence, the brother who travels to the star is younger. Although the result is correct, the explanation is misleading. Some people may falsely assume that the acceleration causes the age difference and that Einstein's general theory of relativity, which deals with noninertial or accelerating reference frames, is required to explain the paradox. But the acceleration incurred by the traveler is incidental, and the paradox can be unraveled by special relativity alone.

A LONG, STRANGE SPACE TRIP

LET US ASSUME that twin brothers, nicknamed the traveler and the homebody, live in Hanover, N.H. They differ in their wanderlust but share a common desire to build a spacecraft that can achieve 0.6 times the speed of light (0.6 *c*). After working on the spacecraft for years, they are ready to launch it, manned by the traveler, toward a star six light-years away.

His vehicle will quickly accelerate to 0.6 *c*. To reach that speed, it would take a little more than 100 days at an acceleration of two g's. Two g's is two times the acceleration of gravity, about what one experiences on a sharp loop on a roller coaster. If, however, the traveler were an electron, he could be accelerated to 0.6 *c* in a tiny fraction of a second. Therefore, the time to reach 0.6 *c* is not central to the argument.

The traveler uses the length-contraction equation of special relativity to measure distance. So the star six light-years away to the homebody appears to be only 4.8 light-years away to the traveler at a speed of 0.6 *c*. Thus, to the traveler, the trip to the star takes only eight years ($^{4.8}/_{0.6}$), whereas the homebody calculates it taking 10 years ($^{6.0}/_{0.6}$). To solve the twin paradox, we need to consider how each twin would view his and the other's clocks during the trip. Let us assume that each twin has a very powerful telescope that permits such observation. Surprisingly, by focusing on the time it takes light to travel between the two, the paradox can be explained.

Both the traveler and homebody set their clocks at zero when the traveler leaves Earth for the star [*see box below*]. When the traveler reaches the star, his clock reads eight years. But when the homebody sees the traveler reach the star, the homebody's clock reads 16 years. Why 16 years? Because, to the homebody, the craft takes 10 years to make it to the star, and the light takes six additional years to come back to Earth, showing the traveler at the star. Hence, viewed through the homebody's telescope, the traveler's clock appears to be running at half the speed of his clock ($^{8}/_{16}$).

As the traveler reaches the star, he reads his clock at eight years as mentioned, but he sees the homebody's clock as it was six years ago (the amount of time it takes for the light from Earth to reach him), or at four years (10 minus 6). So the traveler also views the homebody's clock as running at half the speed of his clock ($^{4}/_{8}$).

FROM TWIN TO YOUNGER BROTHER

ON THE TRIP BACK, the homebody views the traveler's clock going from eight years to 16 years in only four years' time because his clock was at 16 years when he saw the traveler leave the star, and it will be at 20 years when the traveler arrives back home. So the homebody sees the traveler's clock advance eight years in four years of his time; it is now running twice as fast as his clock.

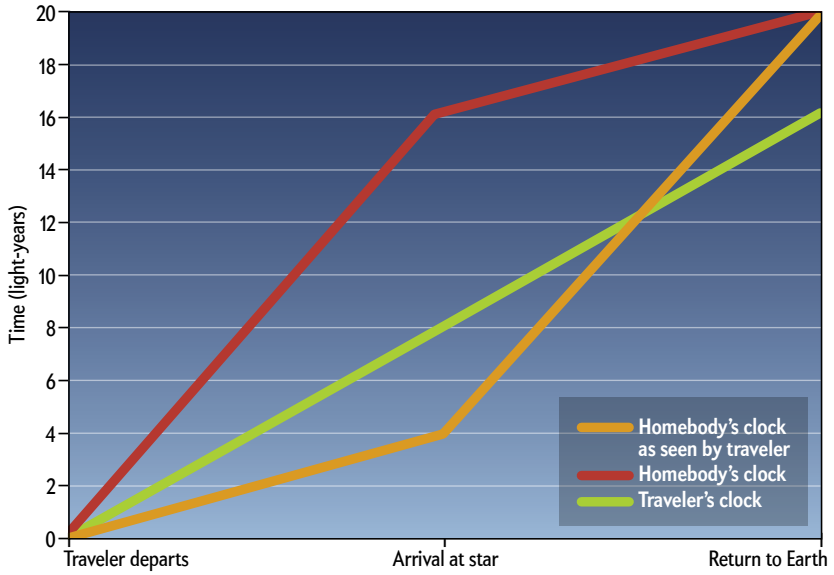
As the traveler returns home, he sees the homebody's clock advance from four to 20 years in eight years of his time. He also sees his brother's clock advancing at twice the speed of his. They both agree, however, that at the end of the trip the traveler's clock reads 16 years and the homebody's 20 years. Therefore, the traveler is four years younger.

The asymmetry in the paradox is that the traveler leaves Earth's reference frame and comes back, whereas the homebody

PARADOX SOLVED

Comparing Clocks

Time passage differs for two twins: a traveler who makes a near-light-speed round-trip to a distant star and a homebody who waits for his return on Earth. At each event—the traveler's departure, his arrival at the star and his return to Earth—both the homebody and the traveler see the same reading on the traveler's clock but different readings on the homebody's clock.



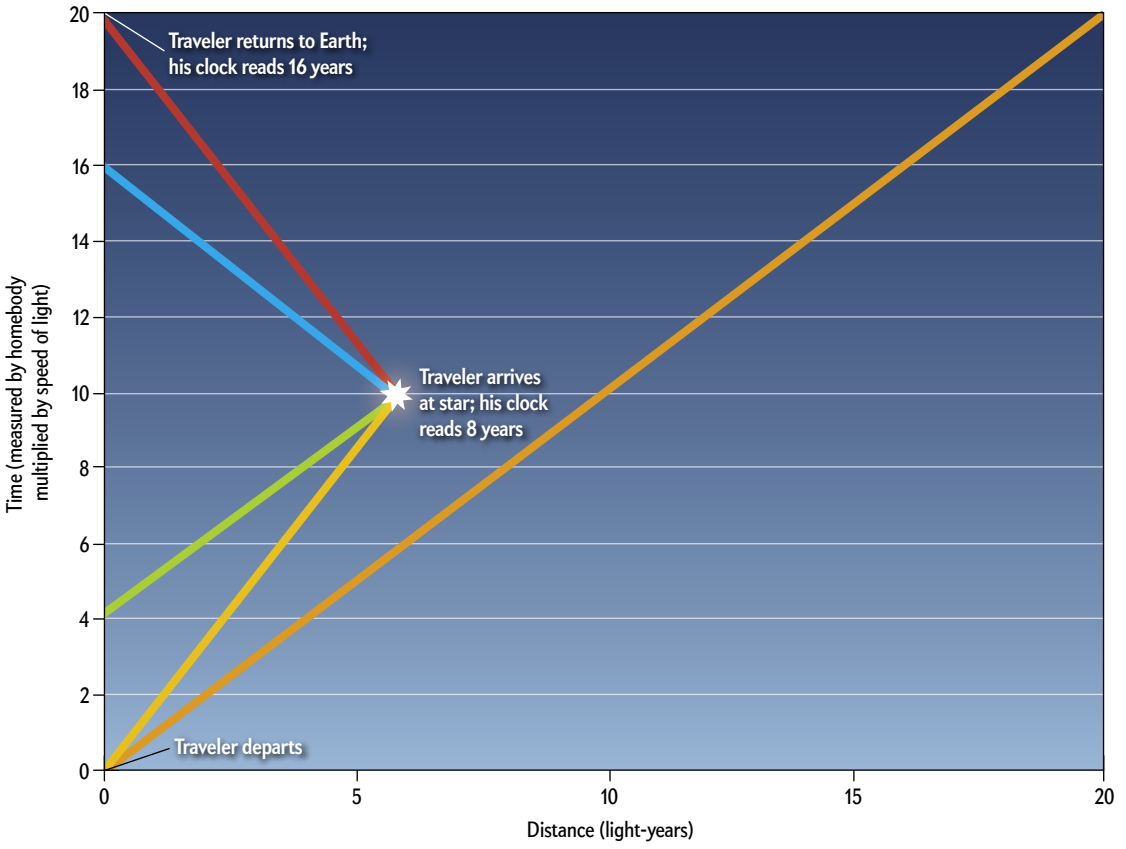
LUCY READING-IKKANDA (*graphs*)

© 2018 Scientific American

Go Far, Younger Man

After a 20-year round-trip to a star at near light speeds, the traveling twin ends up four years younger than the homebody because of the effects of Doppler time dilation.

It takes the traveler eight years to reach the star on the trip out (*yellow*). When he looks back to Earth, he sees the homebody's clock reading only four years as having passed (*green*). But when the homebody sees the traveler at the star, the homebody's clock reads 16 years as having passed (*blue*). On the traveler's arrival back on Earth (*red*), they both agree that the homebody's clock reads 20 years and that the traveler's clock reads 16 years. Hence, the traveler is four years younger. (The orange line shows how light travels over 20 years.)



never leaves Earth. It is also an asymmetry that the traveler and the homebody agree with the reading on the traveler's clock at each event but that they do not agree about the reading on the homebody's clock at each event. The traveler's actions define the events.

The Doppler effect and relativity together explain this effect mathematically at any instant. The reader should also note that the speed at which an observed clock appears to run depends on whether it is traveling away from or toward the observer.

Finally, we should point out that the twin paradox today is more than a theory because its fundamentals have been exhaustively confirmed experimentally. In one such experiment, the lifetime of muon decay verifies the existence of time dilation. Stationary muons have a lifetime of about 2.2 microseconds. When traveling past an observer at 0.9994 c, muons' lifetime stretches to 63.5 microseconds, just as predicted by special relativity. Experiments in which atomic clocks are transported at varying speeds have also produced results that confirm both special relativity and the twin paradox. In the famous Hafele-Keating experiment of 1971, for example, researchers flew cesium-beam atomic clocks around the world onboard commercial airliners—first eastbound and then westbound—and compared their times with stationary clocks at the U.S. Naval Observatory. SA

MORE TO EXPLORE

General Relativity from A to B. Robert Geroch. University of Chicago Press, 1978.
About Time: Einstein's Unfinished Revolution. Paul Davies. Simon & Schuster, 1996.
The Story of Time. Edited by Kristen Lippincott. Merrell, 1999.
Revolution in Time: Clocks and the Making of the Modern World. David S. Landes. Belknap Press, 2000.
The Discovery of Time. Edited by Stuart McCready. Sourcebooks, 2001.
Relativity. Judith Herbst. Twenty-First Century Books, 2007.
Fundamentals of Physics. Tenth edition. David Halliday, Robert Resnick and Jearl Walker. Wiley, 2013.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American

EADWEARD MUYBRIDGE Getty Images

Karen Wright is an award-winning writer and editor based near Seattle. Her work has appeared in numerous national magazines and newspapers and has been syndicated worldwide.

BIOLOGY

TIMES OF OUR LIVES

Whether they're counting minutes, months or years, biological clocks help to keep our brains and bodies running on schedule

By Karen Wright

THE BIOPSYCHOLOGIST JOHN GIBBON called time the "primordial context": a fact of life that has been felt by all organisms in every era. For the morning glory that spreads its petals at dawn, for geese flying south in autumn, for locusts swarming every 17 years and even for lowly slime molds sporing in daily cycles, timing is everything. In human bodies, biological clocks keep track of seconds, minutes, days, months and years. They govern the split-second moves of a tennis serve and account for the trauma of jet lag, monthly surges of menstrual hormones and bouts of wintertime blues. Cellular chronometers may even decide when your time is up. Life ticks, then you die.

IN BRIEF

In the brain, a "stopwatch" can track seconds, minutes and hours.

Another timepiece in the brain, more a clock than a stopwatch, synchronizes many bodily functions with day and night.

This same clock may account for seasonal affective disorder (SAD).

A molecular hourglass that governs the number of times a cell can divide might put a limit on longevity.

© 2018 Scientific American

The pacemakers involved are as different as stopwatches and sundials. Some are accurate and inflexible, others less reliable but subject to conscious control. Some are set by planetary cycles, others by molecular ones. They are essential to the most sophisticated tasks the brain and body perform. And timing mechanisms offer insights into aging and disease. Cancer, Parkinson's disease, seasonal depression and attention-deficit disorder have all been linked to defects in biological clocks.

A VIRTUE OF THE INTERVAL-TIMING STOPWATCH IS ITS FLEXIBILITY. YOU CAN START AND STOP IT AT WILL.

The physiology of these timepieces is not completely understood. But neurologists and other clock researchers have begun to answer some of the most pressing questions raised by human experience in the fourth dimension. Why, for example, a watched pot never boils. Why time flies when you're having fun. Why all-nighters can give you indigestion. Or why people live longer than hamsters. It's only a matter of time before clock studies resolve even more profound quandaries of temporal existence.

THE PSYCHOACTIVE STOPWATCH

IF THIS ARTICLE intrigues you, the time you spend reading it will pass quickly. It'll drag if you get bored. That's a quirk of a "stopwatch" in the brain—the so-called interval timer—that marks time spans of seconds to hours. The interval timer helps you figure out how fast you have to run to catch a baseball. It tells you when to clap to your favorite song. It lets you sense how long you can lounge in bed after the alarm goes off.

Interval timing enlists the higher cognitive powers of the cerebral cortex, the brain center that governs perception, memory and conscious thought. When you approach a yellow traffic light while driving your car, for example, you time how long it has been yellow and compare that with a memory of how long yellow lights usually last. "Then you have to make a judgment about whether to put on the brakes or keep driving," says Stephen M. Rao, now at the Cleveland Clinic Lou Ruvo Center for Brain Health.

Rao's studies with functional magnetic resonance imaging (fMRI) have pointed to the parts of the brain engaged in each of those stages. Inside the fMRI machine, subjects listen to two pairs of tones and decide whether the interval between the second pair is shorter or longer than the interval between the first pair. The brain structures that are involved in the task consume more oxygen than those that are not involved, and the fMRI scan records changes in blood flow and oxygenation once every 250 milliseconds. "When we do this, the very first structures that are activated are the basal ganglia," Rao says.

Long associated with movement, this collection of brain regions has become a prime suspect in the search for the interval-timing mechanism as well. One area of the basal ganglia, the striatum, hosts a population of conspicuously well-connected nerve cells that receive signals from other parts of the brain. The long arms of these striatal cells are covered with between 10,000 and 30,000 spines, each of which gathers information from a different neuron in another locale. If the brain acts like a network, then the striatal spiny neurons are critical nodes. "This is one of only a few places in the brain where you see thousands of neurons converge on a single neuron," says Warren H. Meck of Duke University.

Striatal spiny neurons are central to an interval-timing theory Meck developed with Gibbon, who worked at Columbia University until his death in 2001. The theory posits a collection of neural oscillators in the cerebral cortex: nerve cells firing at different rates, without regard to their neighbors' tempos. In fact, many cortical cells are known to fire at rates between 10 and 40 cycles per second without external provocation. "All these neurons are oscillating on their own schedules," Meck observes, "like people talking in a crowd. None of them are synchronized."

The cortical oscillators connect to the striatum via millions of signal-carrying arms, so the striatal spiny neurons can eavesdrop on all those haphazard "conversations." Then something—a yellow traffic light, say—gets the cortical cells' attention. The stimulation prompts all the neurons in the cortex to fire simultaneously, causing a characteristic spike in electrical output some 300 milliseconds later. This attentional spike acts like a starting gun, after which the cortical cells resume their disorderly oscillations.

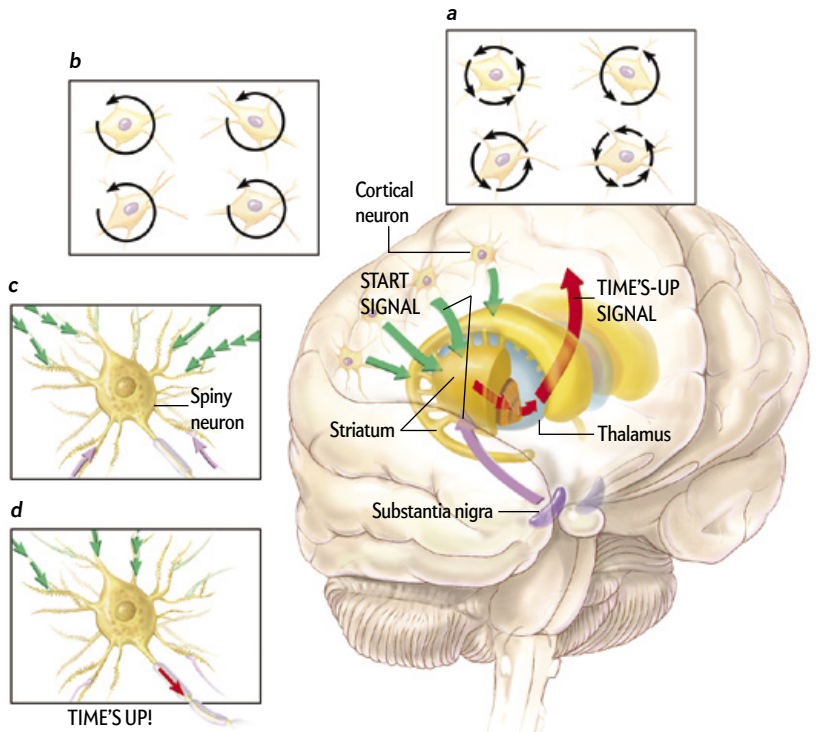
But because they have begun simultaneously, the cycles now make a distinct, reproducible pattern of nerve activation from moment to moment. The spiny neurons monitor those patterns, which help them to "count" elapsed time. At the end of a specified interval—when, for example, the traffic light turns red—a part of the basal ganglia called the substantia nigra sends a burst of the neurotransmitter dopamine to the striatum. The dopamine burst induces the spiny neurons to record the pattern of cortical oscillations they receive at that instant, like a flashbulb exposing the interval's cortical signature on the spiny neurons' film. "There's a unique time stamp for every interval you can imagine," Meck says.

Once a spiny neuron has learned the time stamp of the interval for a given event, subsequent occurrences of the event prompt both the "firing" of the cortical starting gun and a burst of dopamine at the beginning of the interval [*see top illustration in box on opposite page*]. The dopamine burst now tells the spiny neurons to start tracking the patterns of cortical impulses that follow. When the spiny neurons recognize the time stamp marking the end of the interval, they send an electrical pulse from the striatum to another brain center, called the thalamus. The thalamus, in turn, communicates with the cortex, and the higher cognitive functions—such as memory and decision making—take over. Hence, the timing mechanism loops from the cortex to the striatum to the thalamus and back to the cortex again.

© 2018 Scientific American

Clocks in the Brain

Scientists are uncovering the workings of two neural timepieces: an interval timer, which measures intervals lasting up to hours, and a circadian clock, which causes certain body processes to peak and ebb on 24-hour cycles.

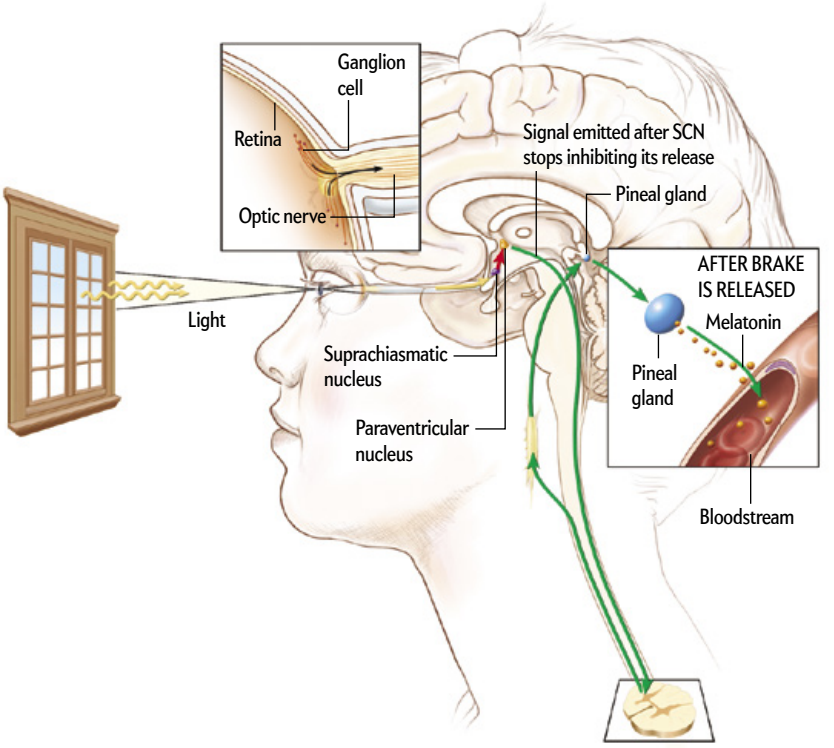


The Interval Timer

According to one model, the onset of an event lasting a familiar amount of time (such as the switching on of a four-second yellow traffic light) activates the "start button" of the interval timer by evoking two brain responses. It induces a particular subset of cortical nerve cells that fire at different rates (*a*) to momentarily act together (*b* and *green arrows on brain*), and it prompts neurons of the substantia nigra to release a burst of the signaling chemical dopamine (*purple arrow*). Both signals impinge on spiny cells of the striatum (*c*), which proceed to monitor the overall patterns of impulses coming from the cortical cells after those neurons resume their various firing rates. Because the cortical cells act in synchrony at the start of the interval, the subsequent patterns occur in the same sequence every time and take a unique form when the end of the familiar interval is reached (*d*). At that point, the striatum sends a time's-up signal (*red arrows*) through other parts of the brain to the decision-making cortex.

The Circadian Clock

Daily cycles of light and dark dictate when many physiological processes that operate on 24-hour cycles will be most and least active. The brain tracks fluctuations in light with the help of ganglion cells in the retina of the eye. A pigment in some of the cells—melanopsin—probably detects light, leading the retinal ganglion cells to send information about its brightness and duration to the suprachiasmatic nucleus (SCN) of the brain. Then the SCN dispatches the information to the parts of the brain and body that control circadian processes. Researchers best understand the events leading the pineal gland to secrete melatonin, sometimes called the sleep hormone (*diagram*). In response to daylight, the SCN emits signals (*red arrow*) that stop another brain region—the paraventricular nucleus—from producing a message that would ultimately result in melatonin's release. After dark, however, the SCN releases the brake, allowing the paraventricular nucleus to relay a "secrete melatonin" signal (*green arrows*) through neurons in the upper spine and the neck to the pineal gland.

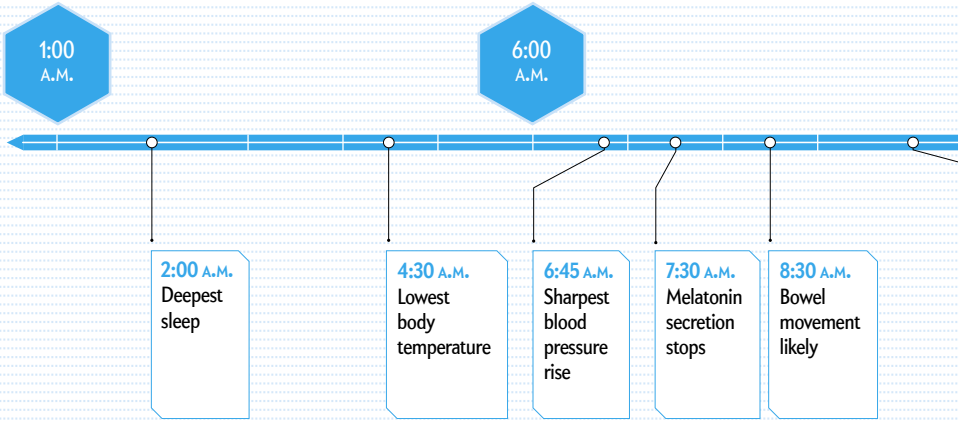


TERESE WINSLOW

© 2018 Scientific American

Rhythm of Life

The circadian clock affects the daily rhythms of many physiological processes. The diagram depicts the circadian patterns typical of someone who rises early in the morning, eats lunch around noon and sleeps at night. Although circadian rhythms tend to be synchronized with cycles of light and dark, other factors—such as ambient temperature, mealtimes, stress and exercise—can influence the timing as well.



If Meck is right and dopamine bursts play an important role in framing a time interval, then diseases and drugs that affect dopamine levels should also disrupt that loop. So far that is what Meck and others have found. Patients with untreated Parkinson's disease, for example, release less dopamine into the striatum, and their clocks run slow. In trials these patients consistently underestimate the duration of time intervals. Marijuana also lowers dopamine availability and slows time. Recreational stimulants such as cocaine and methamphetamine increase the availability of dopamine and make the interval clock speed up, so that time seems to expand. Adrenaline and other stress hormones make the clock speed up, too, which may be why a second can feel like an hour during unpleasant situations. States of deep concentration or extreme emotion may flood the system or bypass it altogether; in such cases, time may seem to stand still or not exist at all. Because an attentional spike initiates the timing process, Meck thinks people with attention-deficit hyperactivity disorder might also have problems gauging the true length of intervals.

The interval clock can also be trained to greater precision. Musicians and athletes know that practice improves their timing; ordinary folk can rely on tricks such as chronometric counting ("one one-thousand") to make up for the mechanism's deficits. Rao forbids his subjects from counting in experiments because it could activate brain centers related to language as well as timing. But counting works, he says—well enough to expose cheaters. "The effect is so dramatic that we can tell whether they're counting or timing based just on the accuracy of their responses."

THE SOMATIC SUNDIAL

ONE OF THE VIRTUES of the interval-timing stopwatch is its flexibility. You can start and stop it at will or ignore it completely. It can work subliminally or submit to conscious control. But it won't win any prizes for accuracy. The precision of interval timers has been found to range from 5 to 60 percent. They don't work too well if you're distracted or tense. And timing errors get worse as an interval gets longer. That is why we rely on cell phones and wristwatches to tell time.

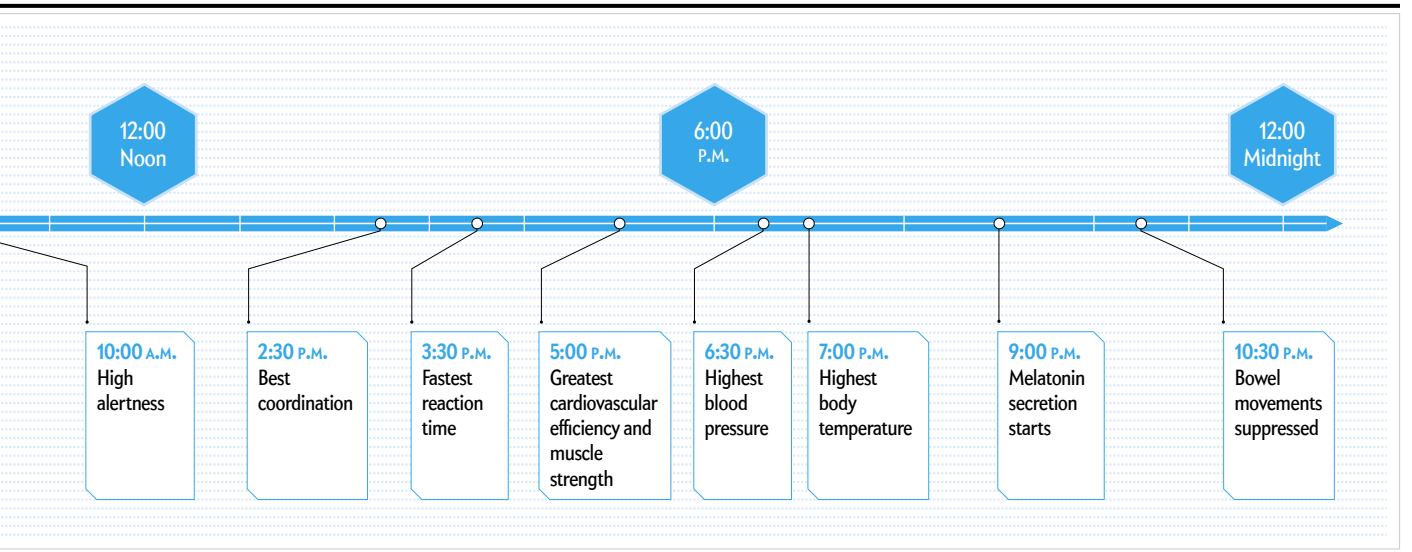
Fortunately, a more rigorous timepiece chimes in at intervals of 24 hours. The circadian clock—from the Latin *circa* ("about") and *diem* ("a day")—tunes our bodies to the cycles of sunlight and darkness that are caused by the earth's rotation. It helps to program the daily habit of sleeping at night and waking in the morning. Its influence extends much further, however. Body temperature regularly peaks in the late afternoon or early evening and bottoms out a few hours before we rise in the morning. Blood pressure typically starts to surge between 6:00 and 7:00 A.M. Secretion of the stress hormone cortisol is 10 to 20 times higher in the morning than at night. Urination and bowel movements are generally suppressed at night and then pick up again in the morning.

The circadian timepiece is more like a clock than a stopwatch because it runs without the need for a stimulus from the external environment. Studies of volunteer cave dwellers and other human guinea pigs have demonstrated that circadian patterns persist even in the absence of daylight, occupational demands and caffeine. Moreover, they are expressed in every cell of the body. Confined to a petri dish under constant lighting, human cells still follow 24-hour cycles of gene activity, hormone secretion and energy production. The cycles are hardwired, and they vary by as little as 1 percent—just minutes a day.

But if light isn't required to establish a circadian cycle, it is needed to synchronize the phase of the hardwired clock with natural day and night cycles. Like an ordinary clock that runs a few minutes slow or fast each day, the circadian clock needs to be continually reset to stay accurate. Neurologists have made great progress in understanding how daylight sets the clock. Two clusters of 10,000 nerve cells in the hypothalamus of the brain have long been considered the clock's locus. Decades of animal studies have demonstrated that these centers, each called a suprachiasmatic nucleus (SCN), drive daily fluctuations in blood pressure, body temperature, activity level and alertness. The SCN also tells the brain's pineal gland when to release melatonin, which promotes sleep in humans and is secreted only at night.

More than 15 years ago researchers

© 2018 Scientific American


12:00 Noon
6:00 P.M.
12:00 Midnight
10:00 A.M. High alertness
2:30 P.M. Best coordination
3:30 P.M. Fastest reaction time
5:00 P.M. Greatest cardiovascular efficiency and muscle strength
6:30 P.M. Highest blood pressure
7:00 P.M. Highest body temperature
9:00 P.M. Melatonin secretion starts
10:30 P.M. Bowel movements suppressed

CAITLIN CHOI; SOURCE: *THE BODY CLOCK GUIDE TO BETTER HEALTH*, BY MICHAEL SMOLENSKY AND LYNNE LAMBERG. HENRY HOLT, 2000

proved that dedicated cells in the retina of the eye transmit information about light levels to the SCN. These cells—a subset of those known as ganglion cells—operate completely independently of the rods and cones that mediate vision, and they are far less responsive to sudden changes in light. That sluggishness befits a circadian system. It would be no good if watching fireworks or going to a movie matinee tripped the mechanism.

The SCN's role in circadian rhythms has been reevaluated in view of other findings. Scientists had assumed that the SCN somehow coordinated all the individual cellular clocks in the body's organs and tissues. Then, in the mid-1990s, researchers discovered four critical genes that govern circadian cycles in flies, mice and humans. These genes turned up not just in the SCN but everywhere else, too. "These clock genes are expressed throughout the whole body, in every tissue," says Joseph Takahashi, now at the University of Texas Southwestern Medical Center. "We didn't expect that."

More recently, researchers at Harvard University found that the expression of more than 1,000 genes in the heart and liver tissue of mice varied in regular 24-hour periods. But the genes that showed these circadian cycles differed in the two tissues, and their expression peaked in the heart at different hours than in the liver. "They're all over the map," says Michael Menaker of the University of Virginia. "Some are peaking at night, some in the morning and some in the daytime."

Menaker has shown that specific feeding schedules can shift the phase of the liver's circadian clock, overriding the light-dark rhythm followed by the SCN. When lab rats that usually ate at will were fed just once a day, for example, peak expression of a clock gene in the liver shifted by 12 hours, whereas the same clock gene in the SCN stayed locked in sync with light schedules. It makes sense that daily rhythms in feeding would affect the liver, given its role in digestion. Researchers think circadian clocks in other organs and tissues may respond to other external cues—including stress, exercise, and temperature changes—that occur regularly every 24 hours.

No one is ready to dethrone the SCN: its authority over body temperature, blood pressure and other core rhythms is still secure. Yet this brain center is no longer thought to rule the peripheral clocks with an iron fist. "We have oscillators in our organs that can function independently of our oscillators in our brain," Takahashi says.

The autonomy of the peripheral clocks makes a phenomenon such as jet lag far more comprehensible. Whereas the interval timer, like a stopwatch, can be reset in an instant, circadian rhythms take days and sometimes weeks to adjust to a sudden shift in day length or time zone. A new schedule of light will slowly reset the SCN clock. But the other clocks may not follow its lead. The body is not only lagging; it's lagging at a dozen different paces.

Jet lag doesn't last, presumably because all those different drummers are able to eventually sync up again. Shift workers, party animals, college students and other night owls face a worse chrononodilemma. They may be leading a kind of physiological double life. Even if they get plenty of shut-eye by day, their core rhythms are still ruled by the SCN—hence, the core functions continue "sleeping" at night. "You can will your sleep cycle earlier or later," says Alfred J. Lewy of the Oregon Health & Science University. "But you can't will your melatonin levels earlier or later, or your cortisol levels, or your body temperature."

Meanwhile their schedules for eating and exercising could be setting their peripheral clocks to entirely different phases from either the sleep-wake cycle or the light-dark cycle. With their body living in so many time zones at once, it's no wonder shift workers have an increased incidence of heart disease, gastrointestinal complaints and, of course, sleep disorders.

A CLOCK FOR ALL SEASONS

JET LAG AND SHIFT WORK are exceptional conditions in which the innate circadian clock is abruptly thrown out of phase with the light-dark cycles or sleep-wake cycles. The same thing can happen every year, albeit less abruptly, when the seasons change. Research shows that although bedtimes may vary, people tend to get up at about the same time in the morning year-round—usually because their dogs, kids, parents or careers demand it. In the winter, at northern latitudes, that

© 2018 Scientific American

means many people wake up two to three hours before the sun makes an appearance. Their sleep-wake cycle is several time zones away from the cues they get from daylight.

The mismatch between day length and daily life could explain the syndrome known as seasonal affective disorder, or SAD. In the U.S., SAD afflicts as many as one in 20 adults with depressive symptoms such as weight gain, apathy and fatigue between October and March. The condition is 10 times more common in the north than the south. Although SAD occurs seasonally, some experts suspect it is actually a circadian problem. Lewy's work suggests that SAD patients would come out of their depression if they could get up at the natural dawn in the winter. In his view, SAD is not so much a pathology as evidence of an adaptive, seasonal rhythm in sleep-wake cycles. "If we adjusted our daily schedules according to the seasons, we might not have seasonal depression," Lewy says. "We got into trouble when we stopped going to bed at dusk and getting up at dawn."

If modern civilization doesn't honor seasonal rhythms, it's partly because human beings are among the least seasonally sensitive creatures around. SAD is nothing compared to the annual cycles other animals go through: hibernation, migration, molting and especially mating, the master metronome to which all other seasonal cycles keep time. It is possible that these seasonal cycles may also be regulated by the circadian clock, which is equipped to keep track of the length of days and nights. Darkness, as detected by the SCN and the pineal gland, prolongs melatonin signals in the long nights of winter and reduces them in the summer. "Hamsters can tell the difference between a 12-hour day, when their gonads don't grow, and a 12-hour-15-minute day, when their gonads do grow," Menaker says.

If seasonal rhythms are so robust in other animals and if humans have the equipment to express them, then how did we ever lose them? "What makes you think we ever had them?" Menaker asks. "We evolved in the tropics." Menaker's point is that many tropical animals don't exhibit dramatic patterns of annual behavior. They don't need them, because the seasons themselves vary so little. Most tropical animals mate without regard to seasons because there is no "best time" to give birth. People, too, are always in heat. As our ancestors gained greater control of their environment over the millennia, seasons probably became an even less significant evolutionary force.

But one aspect of human fertility is cyclical: women and other female primates produce eggs just once a month. The clock that regulates ovulation and menstruation is a well-documented chemical feedback loop that can be manipulated by hormone treatments, exercise and even the presence of other menstruating women. The reason for the specific duration of the menstrual cycle is unknown, though. The fact that it is the same length as the lunar cycle is a coincidence few scientists have bothered to investigate, let alone explain. No convincing link has yet been found between the moon's radiant or gravitational energy and a woman's reproductive hormones. In that regard, the monthly menstrual clock remains a mystery—outdone perhaps only by the ultimate conundrum, mortality.

TIME THE AVENGER

PEOPLE TEND TO EQUATE aging with the diseases of aging—cancer, heart disease, osteoporosis, arthritis and Alzheimer's, to name a few—as if the absence of disease would be enough to confer immortality. Biology suggests otherwise.

Modern humans in developed countries have a life expectancy of more than 70 years. The life expectancy of your average mayfly, in contrast, is a day. Biologists are just beginning to explore why different species have different life expectancies. If your days are numbered, what's doing the counting?

Comparisons within and among animal species, along with research on aging, have challenged many common assumptions about the factors that determine natural life span. The answer cannot lie solely with a species' genetics: worker honeybees, for example, last a few months, whereas queen bees live for years. Still, genetics are important: a single-gene mutation in mice can produce a strain that lives up to 50 percent longer than usual. High metabolic rates can shorten life span, yet many species of birds, which have fast metabolisms,

SEASONAL CLOCKS

Turn, Turn

Most animals experience dramatic seasonal cycles: they migrate, hibernate, mate and molt at specific times of the year. In spring, for example, some 500,000 sandhill cranes make a migratory stopover at the Platte River in Nebraska. As mating season approaches, the testicles of hamsters quadruple in size. These cycles are hardwired: captive ground squirrels continue to hibernate seasonally even when kept in constant temperatures with unvarying periods of light and dark. Likewise, birds in stable lab conditions get restless at migration time and keep molting and fattening in yearly cycles. The only vestige of seasonality in humans may be seasonal affective disorder, commonly referred to as SAD, a yearly bout of depression that strikes some individuals in winter and can be remedied with light therapy—or by sleeping until the sun comes up. —*K.W.*

SANDHILL CRANES

DANITA DELIMONT *Getty Images*

© 2018 Scientific American

live longer than mammals of comparable body size. And big, slow-metabolizing animals do not necessarily outlast the small ones. The life expectancy of a parrot is about the same as a human's. Among dog species, small breeds typically live longer than large ones.

Scientists in search of the limits to human life span have traditionally approached the subject from the cellular level rather than considering whole organisms. So far the closest thing they have to a terminal timepiece is the so-called mitotic clock. The clock keeps track of cell division, or mitosis, the process by which a single cell splits into two. The mitotic clock is like an hourglass in which each grain of sand represents one episode of cell division. Just as there are a finite number of grains in an hourglass, there seems to be a ceiling on how many times normal cells of the human body can divide. In culture they will undergo 60 to 100 mitotic divisions, then call it quits. "All of a sudden they just stop growing," says John Sedivy of Brown University. "They respire, they metabolize, they move, but they will never divide again."

Cultured cells usually reach this state of senescence in a few months. Fortunately, most cells in the body divide much, much more slowly than cultured cells. Eventually—perhaps after 70 years or so—they, too, can get put out to pasture. "What the cells are counting is not chronological time," Sedivy says. "It's the number of cell divisions."

Sedivy has shown that he could squeeze 20 to 30 more cycles out of human fibroblasts by mutating a single gene. This gene encodes a protein called p21, which responds to changes in structures called telomeres that cap the ends of chromosomes. Telomeres are made of the same stuff that genes are: DNA. They consist of thousands of repetitions of a six-base DNA sequence that does not code for any known protein. Each time a cell divides, chunks of its telomeres are lost. Young human embryos have telomeres between 18,000 and 20,000 bases long. By the time senescence kicks in, the telomeres are only 6,000 to 8,000 bases long.

Biologists suspect that cells become senescent when telomeres shrink below some specific length. Titia de Lange of the Rockefeller University has proposed an explanation for this link. In healthy cells, she showed, the chromosome ends are looped back on themselves like a hand tucked in a pocket. The "hand" is the last 100 to 200 bases of the telomere, which are single-stranded, not paired like the rest. With the help of more than a dozen specialized proteins, the single-stranded end is inserted into the double strands upstream for protection.

If telomeres are allowed to shrink enough, "they can no longer do this looping trick," de Lange says. Untucked, a single-stranded telomere end is vulnerable to fusion with other single-stranded ends. The fusion wreaks havoc in a cell by stringing together all the chromosomes. That could be why Sedivy's mutated p21 cells died after they got in their extra rounds of mitosis. Other cells bred to ignore short telomeres have turned cancerous. The job of normal p21 and telomeres themselves may be to stop cells from dividing so much that they die or become malignant. Cellular senescence could actually be prolonging human life rather than spelling its doom. It might be cells' imperfect defense against malignant growth and certain death.

"Our hope is that we'll gain enough information from this reductionist approach to help us understand what's going on in the whole person," de Lange comments.

For now, the link between shortened telomeres and aging is tenuous at best, although you wouldn't know that from some of the outsized claims certain telomere enthusiasts are making. Maria Blasco, a molecular oncologist at the Spanish National Cancer Research Center in Madrid, for example, has developed a $700 blood test that she says may predict life span by measuring the length of a person's telomeres. The test can determine biological age to within a decade, according to one consultant for the company, Life Length, that markets the test.

Other experts point out that telomere length varies so much among individuals that it can't be used as a reliable indicator of biological age. In any case, most cells do not need to keep dividing to do their job—white blood cells that fight infection and sperm precursors being obvious exceptions. Yet many older people do die of simple infections that a younger body could withstand. "Senescence probably has nothing to do with the nervous system," Sedivy says, because most nerve cells do not divide. "On the other hand, it might very well have something to do with the aging of the immune system."

In any case, telomere loss is just one of the numerous insults cells sustain when they divide, says Judith Campisi, a professor at the Buck Institute for Research on Aging in Novato, Calif., and a cell biologist at Lawrence Berkeley National Laboratory. DNA often gets damaged when it is replicated during cell division, so cells that have split many times are more likely to harbor genetic errors than young cells. Genes related to aging in animals and people often code for proteins that prevent or repair those mistakes. And with each mitotic episode, the by-products of copying DNA build up in cell nuclei, complicating subsequent bouts of replication.

"Cell division is very risky business," Campisi observes. So perhaps it is not surprising that the body puts a cap on mitosis. And cheating cell senescence probably wouldn't grant immortality. Once the grains of sand have fallen through the mitotic hourglass, there's no point in turning it over again. SA

MORE TO EXPLORE

The Body Clock Guide to Better Health. Michael Smolensky and Lynne Lamberg. Henry Holt, 2000.

Neuropsychological Mechanisms of Interval Timing Behavior. Matthew S. Matell and Warren H. Meck in *BioEssays*, Vol. 22, No. 1, pages 94–103; January 2000.

The Evolution of Brain Activation during Temporal Processing. Stephen M. Rao, Andrew R. Mayer and Deborah L. Harrington in *Nature Neuroscience*, Vol. 4, No. 3, pages 317–323; March 2001.

The Living Clock: The Orchestrator of Biological Rhythms. John D. Palmer. Oxford University Press, 2002.

scientificamerican.com/magazine/sa

© 2018 Scientific American

Antonio Damasio is university professor, David Dornsife Professor of Neuroscience, and director of the Brain and Creativity Institute at the University of Southern California. He is author of numerous scientific articles and several books, among them *Descartes' Error* and *Self Comes to Mind*.

NEUROLOGY

Remembering WHEN

Several brain structures contribute to "mind time," organizing our experiences into chronologies of remembered events

By Antonio Damasio

IN BRIEF

Researchers understand how the body keeps time through circadian rhythms but not how the brain is able to place events in the proper chronological sequence.

Recent studies suggest that various brain structures, including the hippocampus, basal forebrain and temporal lobe, have some part to play in keeping "mind time."

WE WAKE UP TO TIME, COURTESY OF AN ALARM CLOCK, AND GO THROUGH a day run by time—the meeting, the visitors, the conference call, the luncheon are all set to begin at a particular hour. We can coordinate our own activities with those of others because we all implicitly agree to follow a single system for measuring time, one based on the inexorable rise and fall of daylight. In the course of evolution, humans have developed a biological clock set to this alternating rhythm of light and dark. This clock, located in the brain's hypothalamus, governs what I call body time [see "Times of Our Lives," by Karen Wright, on page 34].

But there is another kind of time altogether. "Mind time" has to do with how we experience the passage of time and how we organize chronology. Despite the steady tick of the clock, duration can seem fast or slow, short or long. And this variability can happen on different scales, from decades, seasons, weeks and hours, down to the tiniest intervals of music—the span of a note or the moment of silence between two notes. We also place events in time, deciding when they occurred, in which order and on what scale, whether that of a lifetime or of a few seconds.

How mind time relates to the biological clock of body time is unknown. It is also not clear whether mind time depends on a single timekeeping device or if our experiences of duration and temporal order rely primarily, or even exclusively, on information processing. If the latter alternative proves to be true, mind time must be determined by the attention we give to events and the emotions we feel when they occur. It must also be influenced by the manner in which we record those events and the inferences we make as we perceive and recall them.

DAVID MUIR Getty Images

© 2018 Scientific American

© 2018 Scientific American

TIME AND MEMORY

I WAS FIRST DRAWN TO the problems of time processing through my work with neurological patients. People who sustain damage to regions of the brain involved in learning and recalling new facts develop major disturbances in their ability to place past events in the correct epoch and sequence. Moreover, these amnesiacs lose the ability to estimate the passage of time accurately at the scale of hours, months, years and decades. Their biological clock, on the other hand, often remains intact, and so can their abilities to sense brief durations lasting a minute or less and to order them properly. At the very least, the experiences of these patients suggest that the processing of time and certain types of memory must share some common neurological pathways.

The association between amnesia and time can be seen most dramatically in cases of permanent brain damage to the hippocampus, a region of the brain important to memory, and to the nearby temporal lobe, the region through which the hippocampus holds a two-way communication with the rest of the cerebral cortex. Damage to the hippocampus prevents the creation of new factual memories. The ability to form memories is an indispensable part of the construction of a sense of our own chronology. We build our time line event by event, and we connect personal happenings to those that occur around us. When the hippocampus is impaired, patients become unable to hold factual memories for longer than about one minute. Patients so afflicted are said to have anterograde amnesia.

Intriguingly, the memories that the hippocampus helps to create are not stored in it. They are distributed in neural networks located in parts of the cerebral cortex (including the temporal lobe) related to the material being recorded: areas dedicated to visual impressions, sounds, tactile information, and so forth. These networks must be activated to both lay down and recall a memory; when they are destroyed, patients cannot recover long-term memories, a condition known as retrograde amnesia. The memories most markedly lost in retrograde amnesia are precisely those that bear a time stamp: recollections of unique events that happened in a particular context on a particular occasion. For instance, the memory of one's wedding bears a time stamp. A different but related kind of recollection—say, that of the concept of marriage—carries no such date with it. The temporal lobe cortex that surrounds the hippocampus is critical for making and recalling such memories.

AMNESIACS LOSE THE ABILITY TO ESTIMATE THE PASSAGE OF TIME ACCURATELY AT THE SCALE OF HOURS, MONTHS, YEARS AND DECADES.

In patients who sustain damage to the temporal lobe cortex, years and even decades of autobiographical memory can be expunged irrevocably. Viral encephalitis, stroke and Alzheimer's disease are among the neurological insults responsible for the most profound impairments.

For one such patient, whom my colleagues and I studied for 25 years, the time gap went almost all the way to the cradle. When this patient was 46, he sustained damage both to the hippocampus and to parts of the temporal lobe. Accordingly, he had both anterograde and retrograde amnesia: he could not form new factual memories, and he could not recall old ones. The patient inhabited a permanent present, unable to remember what happened a minute earlier or 20 years before.

Indeed, he had no sense of time at all. He could not tell us the date, and when asked to guess, his responses were wild—as disparate as 1942 and 2013. He could guess time more accurately if he had access to a window and could approximate it based on light and shadows. But if he was deprived of a watch or a window, morning was no different from afternoon, and night was no different from day; the clock of body time was of no help. This patient could not state his age, either. He could guess, but the guess tended to be wrong.

Two of the few specific things he knew for certain were that he was married and that he was the father of two children. But when did he get married? He could not say. When were the children born? He did not know. He could not place himself in the time line of his family life. He was in fact married, but his wife divorced him more than two decades before. His children had long been married and had children of their own.

TIME STAMPS

HOW THE BRAIN ASSIGNS an event to a specific time and then puts that event in a chronological sequence—or in the case of my patient, fails to do so—is still a mystery. We know only that both the memory of facts and the memory of spatial and temporal relations between those facts are involved. Accordingly, when I was at the University of Iowa, my colleagues Daniel Tranel and Robert Jones and I decided to investigate how an autobiographical time line is established. By looking at people with different kinds of memory impairment, we hoped to identify what region or regions of the brain are required to place memories in the correct epoch.

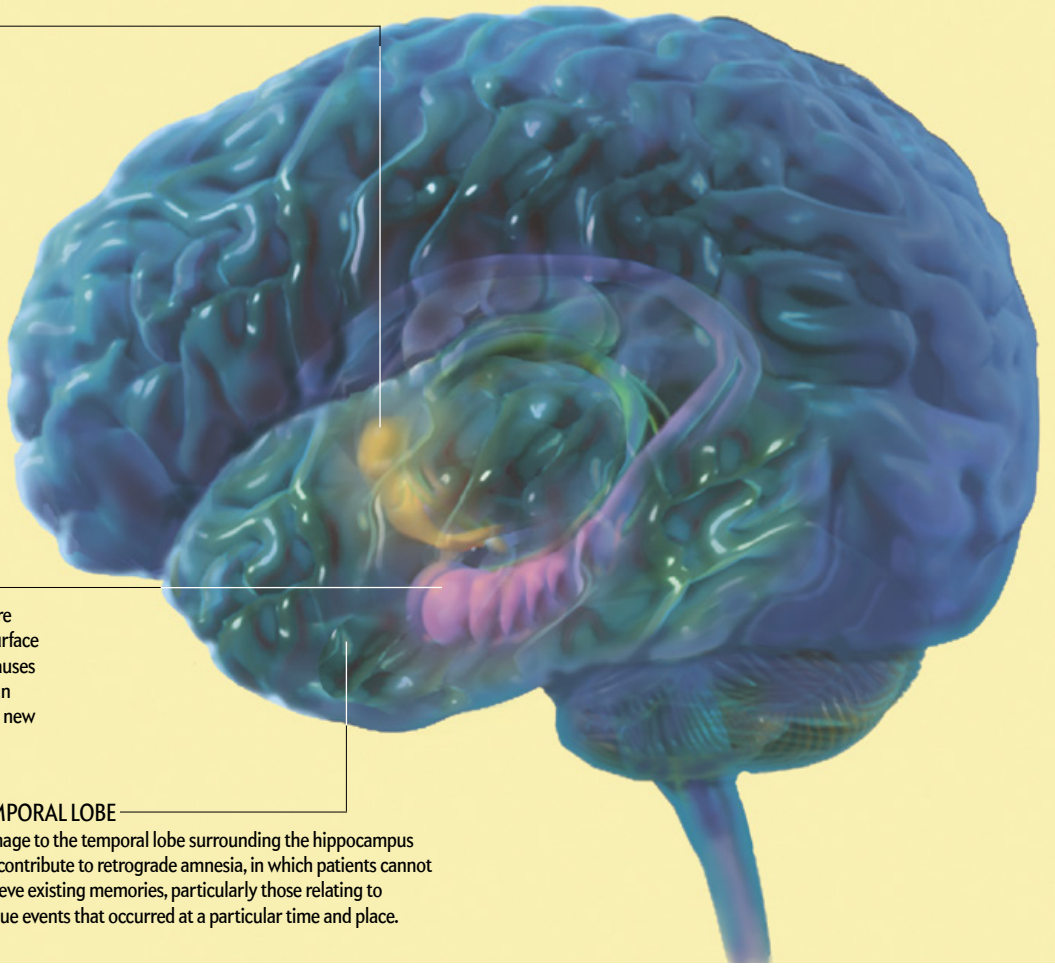
We selected four groups of participants, 20 people in total. The first group consisted of patients with amnesia caused by damage in the temporal lobe. Patients with amnesia caused by damage in the basal forebrain, another area relevant for memory, made up the second set. The third group was composed of patients without amnesia who had damage in places other than the temporal lobe or basal forebrain. We chose as control subjects individuals without neurological disease who had normal memories and who were matched to the patients in terms of age and level of education.

Every participant completed a detailed questionnaire about key events in their life. We asked them about parents, siblings and various relatives, schools, friendships and professional activities. We verified the answers with relatives and records. We also established what the participants remembered of key public events, such as the election of officials, wars and natural disasters, and prominent cultural developments. Then we had each participant place a customized card that described a specific personal or public event on a board that laid out a year-by-year and decade-by-decade time line for the 1900s. For the participants, the situation was an experience similar to playing the board game Life. For the investigators, the setup

© 2018 Scientific American

Finding Time

Studies of brain-damaged patients suggest that structures in the temporal lobe of the brain and in the basal forebrain play important roles in laying down and unearthing information about when events occurred and in what order.



BASAL FOREBRAIN
Injury to this area spares the ability to remember some events but impairs recall of when they happened—indicating that the region plays a role in identifying the chronology of past occurrences.

HIPPOCAMPUS
Damage to this structure (located on the inner surface of the temporal lobe) causes anterograde amnesia: an impaired ability to form new factual memories.

TEMPORAL LOBE
Damage to the temporal lobe surrounding the hippocampus can contribute to retrograde amnesia, in which patients cannot retrieve existing memories, particularly those relating to unique events that occurred at a particular time and place.

CAROL DONNER

permitted a measurement of the accuracy of time placement.

Predictably, the amnesiac patients differed from the controls. Normal individuals were relatively accurate in their time placements: on average, they were wrong by 1.9 years. Amnesiac patients made far more errors, especially those with basal forebrain damage. Although they recalled the event exactly, they were off the mark by an average of 5.2 years. But their recall of events was superior to that of temporal lobe amnesiacs, who were nonetheless more accurate with regard to time stamping—they were off by an average of only 2.9 years.

The results suggest that time stamping and event recall are processes that can be separated. More intriguingly, the outcome indicates that the basal forebrain may be critical in helping to establish the context that allows us to place memories in the right epoch. This notion is in keeping with the clinical observation of basal forebrain patients. Unlike certain of their counterparts with temporal lobe damage, these patients do learn new facts. But they often recall the facts they have just learned in the incorrect order, reconstructing sequences of events in a fictional narrative that can change from occasion to occasion.

BEING LATE FOR CONSCIOUSNESS

MOST OF US do not have to grapple with the large gaps of memory or the chronological confusion that many of my patients do. Yet we all share a strange mental time lag, a phenomenon first brought to light in the 1970s by the late neurophysiologist Benjamin Libet of the University of California, San Francisco. In one experiment, Libet documented a gap between the time an individual was conscious of the decision to flex his finger (and recorded the exact moment of that awareness) and the time his brain waves indicated that a flex was imminent. The brain activity occurred a third of a second before the person consciously decided to move his finger. In another experiment, Libet tested whether a stimulus applied directly to the brain caused any sen-

© 2018 Scientific American

How Hitchcock's *Rope* Stretches Time

The elasticity of time is perhaps best appreciated when we are the spectators of a performance, be it a film, a play, a concert or a lecture. The actual duration of the performance and its mental duration are different things. To illustrate the factors that contribute to this varied experience of time, we can turn to the example of Alfred Hitchcock's 1948 film *Rope*. This technically remarkable work was shot in continuous, unedited 10-minute takes. Few features have been produced in their entirety using this approach. Orson Welles in *Touch of Evil,* Robert Altman in *The Player* and Martin Scorsese in *GoodFellas* employed long, continuous shots but not as consistently as in *Rope.* (In spite of the many plaudits the innovation earned the director, filming proved a nightmare for all concerned, and Hitchcock used the method again only in part of his next film, *Under Capricorn.*)

Hitchcock invented this technique for a sensible reason. He was attempting to depict a story that had been told in a play occurring in continuous time. But he was limited to the amount of film that could be loaded into the camera, which was roughly enough for 10 minutes of action.

Now let us consider how *Rope*'s real time plays in our mind. In an interview with François Truffaut in 1966, Hitchcock stated that the story begins at 7:30 P.M. and terminates at 9:15, 105 minutes later. Yet the film consists of eight reels of 10 minutes each: a total of 80 minutes, including the credits at the beginning and end. Where did the missing 25 minutes go? Do we experience the film as shorter than 105 minutes? Not really. The film never seems shorter than it should, and a viewer has no sense of haste or clipping. On the contrary, for many the film seems longer than its projection time.

I suspect that several aspects account for this alteration of perceived time. First, most of the action takes place in the living room of a penthouse in summer, and the skyline of New York City is visible through a panoramic window. At the beginning of the film the light suggests late afternoon; by the end night has set in. Our daily experience of fading daylight makes us perceive the real-time action as taking long enough to cover the several hours of the coming of night, when in fact, those changes in light are artificially accelerated by Hitchcock.

In the same way, the nature and context of the depicted actions elicit other automatic judgments about duration. After the proverbial Hitchcock murder, which occurs at the beginning of the film's first reel, the story focuses on an elegant dinner party hosted by the two unsavory murderers and attended by the relatives and friends of the victim. The actual time during which food is served is about two reels. Yet viewers attribute more time to that sequence because we know that neither the hosts nor the guests, who look cool, polite and unhurried, would swallow dinner at such breakneck speed. When the action later splits—some guests converse in the living room in front of the camera, while others repair to the dining room to look at rare books—we sensibly attribute a longer duration to this offscreen episode than the few minutes it takes up in the actual film.

Another factor may also contribute to the deceleration of time. There are no jump cuts within each 10-minute reel; the camera glides slowly toward and away from each character. Yet to join each segment to the next, Hitchcock finished most takes with a close-up on an object. In most instances, the camera moves to the back of an actor wearing a dark suit, and the screen goes black for a few seconds;

© 2018 Scientific American

***Rope*'s skyline light fades more quickly than in real life, but viewers attribute real time to the coming of night. They therefore experience time as passing more slowly than it does in the film.**

EVERETT COLLECTION

the next take begins as the camera pulls away from the actor's back. Although the interruption is brief and is not meant to signal a time break, it may nonetheless contribute to the elongation of time because we are used to interpreting breaks in the continuity of visual perception as a lapse in the continuity of time. Film-editing devices such as the dissolve and the fade often cause spectators to infer that time has passed between the preceding shot and the following one. In *Rope,* each of the seven breaks delays real time by a fraction of a second. But cumulatively for some viewers, the breaks may suggest that more time has passed.

The emotional content of the material may also extend time. When we are uncomfortable or worried, we often experience time more slowly because we focus on negative images associated with our anxiety. Studies in my laboratory suggest that the brain generates images at faster rates when we are experiencing positive emotions (perhaps this is why time flies when we're having fun) and reduces the rate of image making during negative emotions. On a recent flight with heavy turbulence, for instance, I experienced the passage of time as achingly slow, perhaps because my attention was directed to the discomfort of the experience. The unpleasantness of the situation in *Rope* may similarly conspire to stretch time.

Rope provides a noticeable discrepancy between real time and the audience's perception of time. In so doing, it illustrates how the experience of duration is a construct. It is based on factors as varied as the content of the events being perceived, the emotional reactions those events provoke and the way in which images are presented to us, as well as the conscious and unconscious inferences that accompany the film's images. —*A.D.*

sation in some patients undergoing brain surgery, who were awake, as most patients are in such operations. He found that a mild electrical charge to the cortex produced a tingling in the patient's hand—a full half a second after the stimulus was applied.

Although the interpretation of those experiments, and of others in the field of consciousness studies, is entangled in controversy, one general fact emerged from Libet's work. It is apparent that a lag exists between the beginning of the neural events leading to awareness and the moment one actually experiences the consequence of those events.

This finding may be shocking at first glance, and yet the reasons for the delay are fairly obvious. It takes time for the physical changes that constitute an event to impinge on the body and to modify the sensory detectors of an organ such as the retina. It takes time for the resulting electrochemical modifications to be transmitted as signals to the central nervous system. It takes time to generate a neural pattern in the brain's sensory maps. Finally, it takes time to relate the neural map of the event and the mental image arising from it to the neural map and image of the self—that is, the notion of who we are—the last and critical step without which the event will never become conscious.

We are talking about nothing more than mere milliseconds, but there is a delay nonetheless. This situation is so strange that the reader may well wonder why we are not aware of this delay. One attractive explanation is that because we have similar brains and they work similarly, we are all hopelessly late for consciousness and no one notices it. But perhaps other reasons apply. At the microtemporal level, the brain manages to "antedate" some events so that delayed processes can appear less delayed and differently delayed processes can appear to have similar delays.

This possibility, which Libet contemplated, may explain why we maintain the illusion of continuity of time and space when our eyes move quickly from one target to another. We notice neither the blur that attends the eye movement nor the time it takes to get the eyes from one place to the other. Patrick Haggard and John C. Rothwell of University College London suggest that the brain predates the perception of the target by as much as 120 milliseconds, thereby giving us the perception of seamless viewing.

The brain's ability to edit visual experiences and to impart a sense of volition after neurons have already acted is an indication of its exquisite sensitivity to time. Although our understanding of mind time is incomplete, we are gradually coming to know more about why we experience time so variably and about what the brain needs to create a time line. SA

MORE TO EXPLORE

Time and the Observer: The Where and When of Consciousness in the Brain. Daniel C. Dennett and Marcel Kinsbourne in *Behavioral and Brain Sciences,* Vol. 15, No. 2, pages 183–247; 1992.

The Influence of Affective Factors on Time Perception. Alessandro Angrilli, Paolo Cherubini, Antonella Pavese and Sara Manfredini in *Perception and Psychophysics,* Vol. 59, No. 6, pages 972–982; August 1997.

From Physical Time to the First and Second Moments of Psychological Time. Simon Grondin in *Psychological Bulletin,* Vol. 127, No. 1, pages 22–44; January 2001.

Illusory Perceptions of Space and Time Preserve Cross-Saccadic Perceptual Continuity. Kielan Yarrow, Patrick Haggard, Ron Heal, Peter Brown and John C. Rothwell in *Nature,* Vol. 414, pages 302–305; November 15, 2001.

Time Perception: Brain Time or Event Time? Alan Johnston and Shin'ya Nishida in *Current Biology,* Vol. 11, No. 11, pages R427–R430; 2001.

scientificamerican.com/magazine/sa

© 2018 Scientific American

ANTHROPOLOGY

CLOCKING CULTURES

What is time? The answer varies from society to society

By the Editors

IN BRIEF

The way various cultures keep time reflects their priorities and even the way they view the world. Despite the near universal use of clocks and calendars, different societies march to different beats.

In perceiving time, cultures tend to emphasize the past, present and future differently. For example, the followers of Wahhabism—the strict form of Islam that prevails in Saudi Arabia—are intent on replicating a romanticized vision of the past.

SHOW UP AN HOUR LATE IN BRAZIL, AND NO ONE bats an eyelash. But keep someone in Switzerland waiting for five or 10 minutes, and you have some explaining to do. Time is elastic in many cultures but snaps taut in others. Indeed, the way members of a culture perceive and use time reflects their society's priorities and even their own worldview.

Social scientists have recorded wide differences in the pace of life in various countries and in how societies view time—whether as an arrow piercing the future or as a revolving wheel in which past, present and future cycle endlessly. Some cultures conflate time and space: the Australian Aborigines' concept of the "Dreamtime" encompasses not only a creation myth but a method of finding their way around the countryside. Interestingly, however, some views of time—such as the idea that it is acceptable for a more powerful person to keep someone of lower status waiting—cut across cultural differences and seem to be found universally.

The study of time and society can be divided into the pragmatic and the cosmolog-

GETTY IMAGES

© 2018 Scientific American

ical. On the practical side, in the 1950s anthropologist Edward T. Hall, Jr., wrote that the rules of social time constitute a "silent language" for a given culture. The rules might not always be made explicit, he stated, but "they exist in the air.... They are either familiar and comfortable or unfamiliar and wrong."

In 1955 he described in *Scientific American* how differing perceptions of time can lead to misunderstandings between people from separate cultures. "An ambassador who has been kept waiting for more than half an hour by a foreign visitor needs to understand that if his visitor 'just mutters an apology' this is not necessarily an insult," Hall wrote. "The time system in the foreign country may be composed of different basic units, so that the visitor is not as late as he may appear to us. You must know the time system of the country to know at what point apologies are really due.... Different cultures simply place different values on the time units."

Most cultures around the world now have clocks and calendars, uniting the majority of the globe in the same general rhythm of time. But that doesn't mean we all march to the same beat. Some people feel so rushed by the pace of modern life that they are fighting back with "slow food," while in other societies, people feel little pressure to "manage" their time.

"One of the beauties of studying time is that it's a wonderful window on culture," says Robert V. Levine, a social psychologist at California State University, Fresno. "You get answers on what cultures value and believe in. You get a really good idea of what's important to people."

Levine and his colleagues have conducted so-called pace-of-life studies in 31 countries. In *A Geography of Time,* first published in 1997, Levine describes how he ranked the countries by using three measures: walking speed on urban sidewalks, how quickly postal clerks could fulfill a request for a common stamp, and the accuracy of public clocks. Based on these variables, he concluded that the five fastest-paced countries are Switzerland, Ireland, Germany, Japan and Italy; the five slowest are Syria, El Salvador, Brazil, Indonesia and Mexico. The U.S., at 16th, ranks near the middle.

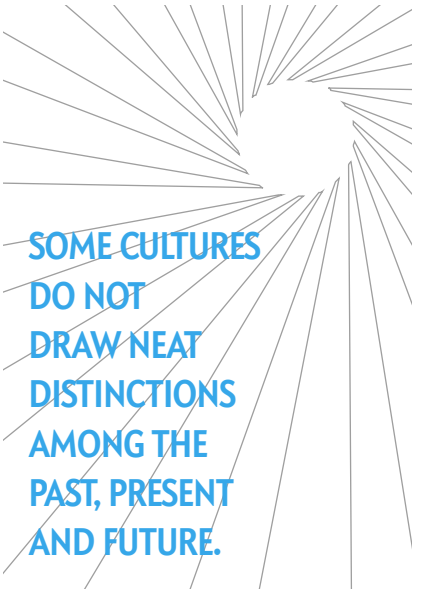
Kevin K. Birth, an anthropologist at Queens College, has examined time perceptions in Trinidad. Birth's 1999 book, *Any Time Is Trinidad Time: Social Meanings and Temporal Consciousness,* refers to a commonly used phrase to excuse lateness. In that country, Birth observes, "if you have a meeting at 6:00 at night, people show up at 6:45 or 7:00 and say, 'Any time is Trinidad time.'" When it comes to business, however, that loose approach to timeliness works only for the people with power. A boss can show up late and toss off "any time is Trinidad time," but underlings are expected to be more punctual. For them, the saying goes, "time is time." Birth adds that the tie between power and waiting time is true for many other cultures as well.

The nebulous nature of time can make it difficult for anthropologists and social psychologists to study. "You can't simply go into a society, walk up to some poor soul and say, 'Tell me about your notions of time,'" Birth says. "People don't really have an answer to that. You have to come up with other ways to find out."

Birth attempted to get at how Trinidadians value time by exploring how closely their society links time and money. He surveyed rural residents and found that farmers—whose days are dictated by natural events, such as sunrise—did not recognize the phrases "time is money," "budget your time" or "time management," even though they had satellite TV and were familiar with Western popular culture. But tailors in the same areas were aware of such notions. Birth concluded that wage work altered the tailors' views of time. "The ideas of associating time with money are not found globally," he says, "but are attached to your job and the people you work with."

How people deal with time on a day-to-day basis often has nothing to do with how they conceive of time as an abstract entity. "There's often a disjunction between how a culture views the mythology of time and how [people] think about time in their daily lives," Birth asserts. "We don't think of Stephen Hawking's theories as we go about our daily lives."

Some cultures do not draw neat distinctions among the past, present and future. Australian Aborigines, for instance, believe that their ancestors crawled out of the earth during the Dreamtime. The ancestors "sang" the world into existence as they moved about naming each feature and living thing, which brought them into being. Even today an entity does not exist unless an Aborigine "sings" it.



Ziauddin Sardar, a British Muslim author and critic, has written about time and Islamic cultures, particularly the fundamentalist sect Wahhabism. Muslims "always carry the past with them," claims Sardar, who is editor of the journal *Critical Muslim* and former chairperson of the Muslim Institute in London. "In Islam, time is a tapestry incorporating the past, present and future. The past is ever present." The followers of Wahhabism, which is widely practiced in Saudi Arabia and by members of al Qaeda, seek to re-create the idyllic days of the prophet Muhammad's life. "The worldly future dimension has been suppressed" by them, Sardar says. "They have romanticized a particular vision of the past. All they are doing is trying to replicate that past."

Sardar asserts that the West has "colonized" time by spreading the expectation that life should become better as time passes: "If you colonize time, you also colonize the future. If you think of time as an arrow, of course you think of the future as progress, going in one direction. But different people may desire different futures." SA

MORE TO EXPLORE

A Geography of Time: The Temporal Misadventures of a Social Psychologist. Robert V. Levine. Basic Books, 1998.

Any Time Is Trinidad Time: Social Meanings and Temporal Consciousness. Kevin K. Birth. University Press of Florida, 1999.

scientificamerican.com/magazine/sa

© 2018 Scientific American

© 2018 Scientific American

William J. H. Andrewes is a museum consultant and maker of precision sundials (www.hodinkee.com/features/will-andrewes-feature). Specializing in the field of time measurement for four decades, he has worked at Eton College, the Royal Observatory Greenwich, The Time Museum and Harvard University. He organized the Longitude Symposium at Harvard (1993), edited *The Quest for Longitude*, and was co-author with Dava Sobel of *The Illustrated Longitude*. For his contributions to horology, he was awarded the Harrison Gold Medal in 2007.

INVENTION

A CHRONICLE OF TIMEKEEPING

Our conception of time depends on the way we measure it

By William J. H. Andrewes

HUMANKIND'S EFFORTS TO TELL TIME HAVE HELPED DRIVE THE EVOLUTION OF our technology and science throughout history. The need to gauge the divisions of the day and night led the ancient Egyptians, Greeks and Romans to create sundials, water clocks and other early chronometric tools. Western Europeans adopted these technologies, but by the 13th century, demand for a dependable timekeeping instrument led medieval artisans to invent the mechanical clock. Although this new device satisfied the requirements of monastic and urban communities, it was too inaccurate and unreliable for scientific application until the pendulum was employed to govern its operation. The precision timekeepers that were subsequently developed resolved the critical problem of finding a ship's position at sea and went on to play key roles in the industrial revolution and the advance of Western civilization.

Today highly accurate timekeeping instruments set the beat for most of our electronic devices. Nearly all computers, for example, contain a quartz-crystal clock to regulate their operation. Moreover, not only do time signals beamed down from Global Positioning System satellites calibrate the functions of precision navigation equipment, they do so as well for cell phones, instant stock-trading systems and nationwide power-distribution grids. So integral have these time-based technologies become to our day-to-day lives that we recognize our dependency on them only when they fail to work.

RECKONING DATES

ACCORDING TO archaeological evidence, the Babylonians, Egyptians and other early civilizations began to measure time at least 5,000 years ago, introducing calendars to organize and coordinate communal activities and public events, to schedule the shipment of goods and, in particular, to regulate cycles of planting and harvesting. They based their calendars on three natural cycles: the solar day, marked by the successive periods of light and darkness as the earth rotates on its axis; the lunar month, following the phases of the moon as it orbits the earth; and the solar year, defined by the changing seasons that accompany our planet's revolution around the sun.

Before the invention of artificial light, the moon had greater social impact. And for those living near the equator in particular, its waxing and waning was more conspicuous than

IN BRIEF

Devices for measuring time have been around for at least 5,000 years and today are essential for coordinating the operation of everything from cell phones to power-distribution grids.

The earliest mechanical clocks, invented just before 1300, told time by striking a bell.

By the early 1500s inventors had developed spring-driven mechanical clocks that were small enough to be portable.

Modern quartz and atomic clocks have made it possible to keep extremely accurate time, which has opened the door to new applications.

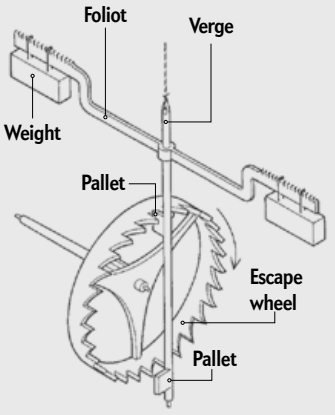
ALAMY (*watch face*); SCIENTIFIC AMERICAN (*digital question marks*)

© 2018 Scientific American

Development of Early Mechanical Clockworks

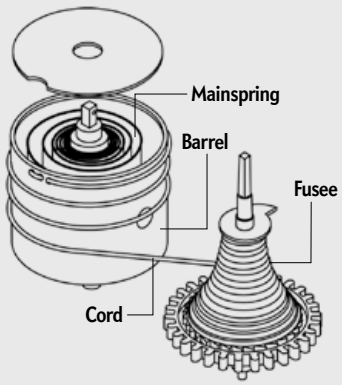
Verge and Foliot Escapement

The innovative component of the first mechanical clocks (circa 1300) was the escapement, a device that both controlled the escape wheel's rotation and transmitted the power needed to sustain the motion of the oscillator, which in turn regulated the speed at which the timekeeper operated. The escape wheel, a sawtoothed crown, is driven by a gear train powered by a weighted cord wound around the axle. The clockwise rotation of the escape wheel is obstructed by two pallets protruding from a vertical shaft, called a verge, which carries a bar known as a foliot. When the top pallet checks the escape wheel's rotation (causing a "tick"), the engaged wheel tooth gradually forces the pallet back until it is free to escape. The wheel's movement, however, is stopped almost immediately when the lower pallet arrests another tooth (causing a "tock") and then pushes the verge in the opposite direction. Driven by the escape wheel, the to-and-fro oscillation of the verge and foliot continues until the cord fully unwinds. The rate at which the mechanism operates can be adjusted by moving the weights on the foliot arms out (for slower) and in (for faster).



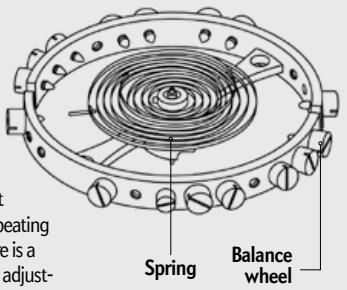
Fusee

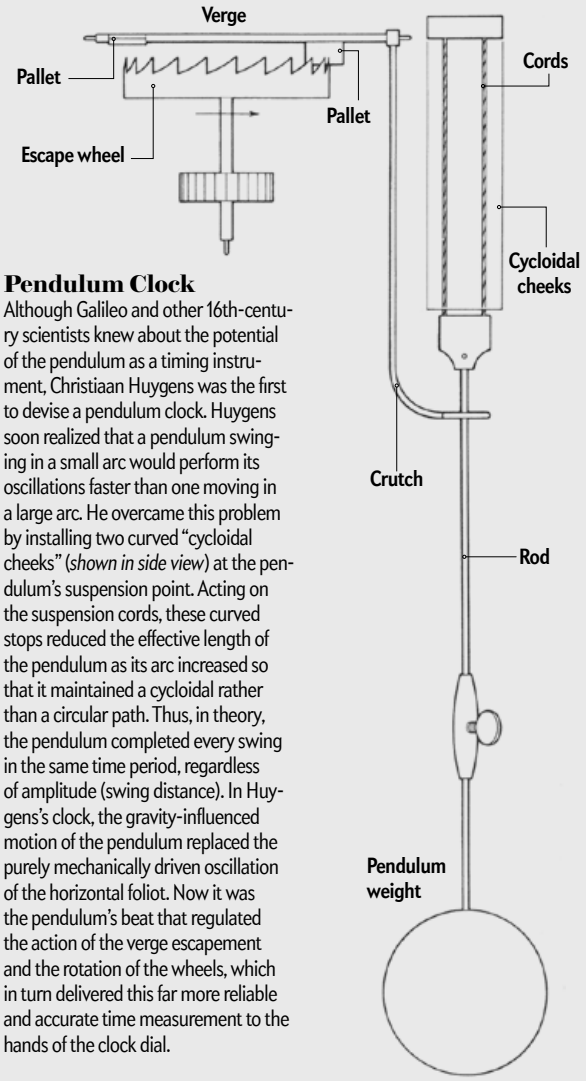
The use of coiled springs as the motive force for timekeepers was made practical by the invention of the fusee in the early to mid-1400s. Although a spring is a compact power source, its force varies, increasing as it is wound more tightly. The fusee, a cone-shaped grooved pulley, was devised to compensate for the variable strength of a timekeeper's mainspring. The barrel, which houses the spring, is connected to the fusee by a cord or chain. When the mainspring is fully wound, the cord pulls on the narrow end of the fusee, where a short torque arm produces relatively little leverage. As the clock runs, the cord is gradually drawn back onto the barrel. To compensate for the mainspring's diminishing strength, the cord's spiral track on the fusee increases in diameter. Thus, the force delivered to the gear wheels of the timekeeper remains constant despite the changing tension of its mainspring.



Spiral Balance Spring

In 1675 Huygens invented the spiral balance spring. Just as gravity controls the swinging oscillation of a pendulum in a clock, this spring regulates the rotary oscillation of a balance wheel in portable timepieces. A balance wheel is a rotor that spins one way and then the other, repeating the cycle over and over. Depicted here is a modern version, finely balanced with adjustable timing screws.



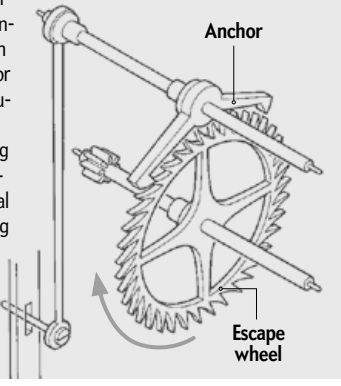


Pendulum Clock

Although Galileo and other 16th-century scientists knew about the potential of the pendulum as a timing instrument, Christiaan Huygens was the first to devise a pendulum clock. Huygens soon realized that a pendulum swinging in a small arc would perform its oscillations faster than one moving in a large arc. He overcame this problem by installing two curved "cycloidal cheeks" (*shown in side view*) at the pendulum's suspension point. Acting on the suspension cords, these curved stops reduced the effective length of the pendulum as its arc increased so that it maintained a cycloidal rather than a circular path. Thus, in theory, the pendulum completed every swing in the same time period, regardless of amplitude (swing distance). In Huygens's clock, the gravity-influenced motion of the pendulum replaced the purely mechanically driven oscillation of the horizontal foliot. Now it was the pendulum's beat that regulated the action of the verge escapement and the rotation of the wheels, which in turn delivered this far more reliable and accurate time measurement to the hands of the clock dial.

Anchor Escapement

Developed around 1670 in England, the anchor escapement is a lever-based device shaped like a ship's anchor. The motion of a pendulum rocks the anchor so that it catches and then releases each tooth of the escape wheel, which, in turn, provides an impulse to maintain the pendulum's oscillation. Unlike the verge escapement used in early pendulum clocks, the anchor escapement permitted the pendulum to travel in such a small arc that maintaining a cycloidal swing path became unnecessary. Moreover, this invention made practical the use of a long, seconds-beating pendulum and thus led to the development of a new, floor-standing case design, which became known as the long-case, or grandfather, clock.



DAVID PENNEY

© 2018 Scientific American

the passing of the seasons. Hence, the calendars developed at the lower latitudes were influenced more by the lunar cycle than by the solar year. In more northern climes, however, where seasonal agriculture was important, the solar year became more crucial. As the Roman Empire expanded northward, it organized its calendar for the most part around the solar year. Today's Gregorian calendar derives from the Babylonian, Egyptian, Jewish and Roman calendars.

The Egyptians formulated a civil calendar having 12 months of 30 days, with five days added to approximate the solar year. Each period of 10 days was marked by the appearance of special star groups (constellations) called decans. At the rise of the star Sirius just before sunrise, which occurred around the all-important annual flooding of the Nile, 12 decans could be seen spanning the heavens. The cosmic significance the Egyptians placed in the 12 decans led them to develop a system in which each interval of darkness (and later, each interval of daylight) was divided into a dozen equal parts. These periods became known as temporal hours because their duration varied according to the changing length of days and nights with the passing of the seasons. Summer hours were long, winter ones short; only at the spring and autumn equinoxes were the hours of daylight and darkness equal. Temporal hours, which were adopted by the Greeks and then the Romans (who spread them throughout Europe), remained in use for more than 2,500 years.

Ingenious inventors devised sundials, which indicate time by the length or direction of the sun's shadow, to track temporal hours during the day. The sundial's nocturnal counterpart, the water clock, was designed to measure temporal hours at night. One of the first water clocks was a basin with a small hole near the bottom through which the water dripped out. The falling water level denoted the passing hour as it dipped below hour lines inscribed on the inner surface. Although these devices performed satisfactorily around the Mediterranean, they could not always be depended on in the cloudy and often freezing weather of northern Europe.

THE PULSE OF TIME

THE EARLIEST RECORDED weight-driven mechanical clock was installed in 1283 at Dunstable Priory in Bedfordshire, England. That the Roman Catholic Church should have played a major role in the invention and development of clock technology is not surprising: the strict observance of prayer times by monastic orders occasioned the need for a more reliable instrument of time measurement. Further, the Church not only controlled education but also possessed the wherewithal to employ the most skillful craftsmen. Additionally, the growth of urban mercantile populations in Europe during the second half of the 13th century created demand for improved timekeeping devices. By 1300 artisans were building clocks for churches and cathedrals in France and Italy. Because the initial examples indicated the time by striking a bell (thereby alerting the surrounding community to its daily duties), the name for this new machine was adopted from the Latin word for "bell," *clocca*.

The revolutionary aspect of this new timekeeper was neither the descending weight that provided its motive force nor the gear wheels (which had been around for at least 1,300 years) that transferred the power; it was the part called the escapement. This device controlled the wheels' rotation and transmitted the power required to maintain the motion of the oscillator, the part that regulated the speed at which the timekeeper operated [*for an explanation of early clockworks, see box on opposite page*]. The inventor of the clock escapement is unknown.

UNIFORM HOURS

THE MECHANICAL CLOCK could be adjusted to maintain temporal hours, but it was naturally suited to keeping equal ones. With uniform hours, however, arose the question of when to begin counting them, and so, in the early 14th century, a number of systems evolved. The schemes that divided the day into 24 equal parts varied according to the start of the count: Italian hours began at sunset, Babylonian hours at sunrise, astronomical hours at midday and "great clock" hours (used for some large public clocks in Germany) at midnight. Eventually these and competing systems were superseded by "small clock," or French, hours, which split the day, as we currently do, into two 12-hour periods commencing at midnight.

During the 1580s clockmakers received commissions for timekeepers showing minutes and seconds, but their mechanisms were insufficiently accurate for these fractions to be included on dials until the 1660s, when the pendulum clock was developed. Minutes and seconds derive from the sexagesimal partitions of the degree introduced by Babylonian astronomers. The word "minute" has its origins in the Latin *prima minuta,* the first small division; "second" comes from *secunda minuta,* the second small division. The sectioning of the day into 24 hours and of hours and minutes into 60 parts became so well established in Western culture that all efforts to change this arrangement failed. The most notable attempt took place in revolutionary France in the 1790s, when the government adopted the decimal system. Although the French successfully introduced the meter, liter and other base-10 measures, the bid to break the day into 10 hours, each consisting of 100 minutes split into 100 seconds, lasted only 16 months.

PORTABLE CLOCKS

FOR CENTURIES after the invention of the mechanical clock, the periodic tolling of the bell in the town church or clock tower was enough to demarcate the day for most people. But by the 15th century a growing number of clocks were being made for domestic use. Those who could afford the luxury of owning a clock found it convenient to have one that could be moved from place to place. Innovators accomplished portability by replacing the weight with a coiled spring. The tension of a spring, however, is greater after it is wound. The contrivance that overcame this problem, known as a fusee (from *fusus,* the Latin term for "spindle"), was invented by an unknown mechanical genius probably between 1400 and 1450 [*see middle left illustration on opposite page*]. This cone-shaped device was connected by a cord to the barrel housing the spring: when the clock was wound, drawing the cord from the barrel onto the fusee, the diminishing diameter of the spiral of the fusee compensated for the increasing pull of the spring. Thus, the fusee equalized the force of the spring on the wheels of the timekeeper.

The importance of the fusee should not be underestimated: it made possible the development of the portable clock as well as the subsequent evolution of the pocket watch. Many high-grade, spring-driven timepieces, such as marine chronometers, continued to incorporate this device until after World War II.

© 2018 Scientific American

INNOVATIVE CLOCKWORKS

IN THE 16TH CENTURY Danish astronomer Tycho Brahe and his contemporaries tried to use clocks for scientific purposes, yet even the best ones were still too unreliable. Astronomers in particular needed a better tool for timing the transit of stars and thereby creating more accurate maps of the heavens. The pendulum proved to be the key to boosting the accuracy and dependability of timekeepers. Galileo, the Italian physicist and astronomer, and others before him experimented with pendulums, but a 27-year-old Dutch astronomer and mathematician named Christiaan Huygens devised the first pendulum clock on Christmas Day in 1656. Huygens recognized the commercial as well as the scientific significance of his invention immediately, and within six months a local maker in the Hague had been granted a license to manufacture pendulum clocks.

Huygens saw that a pendulum traversing a circular arc completed small oscillations faster than large ones. Therefore, any variation in the extent of the pendulum's swing would cause the clock to gain or lose time. Realizing that maintaining a constant amplitude (amount of travel) from swing to swing was impossible, Huygens devised a pendulum suspension that caused the bob to move in a cycloid-shaped arc rather than a circular one. In theory, this enabled it to oscillate in the same time regardless of its amplitude [*see top right illustration in box on page 52*]. Pendulum clocks were about 100 times as accurate as their predecessors, reducing a typical gain or loss of 15 minutes a day to about a minute a week. News of the invention spread rapidly, and by 1660 English and French artisans were developing their own versions of this new timekeeper.

The advent of the pendulum not only heightened demand for clocks but also resulted in their development as furniture. National styles soon began to emerge: English makers designed the case to fit around the clock movement; in contrast, the French placed greater emphasis on the shape and decoration of the case. Huygens, however, had little interest in these fashions, devoting much of his time to improving the device both for astronomical use and for solving the problem of finding longitude at sea.

In 1675 Huygens devised another fundamental improvement, the spiral balance spring. Just as gravity controls the swinging oscillation of a pendulum in clocks, this spring regulates the rotary oscillation of a balance wheel in portable timepieces. A balance wheel is a finely balanced disk that rotates fully one way and then the other, repeating the cycle over and over [*see bottom left illustration in box on page 52*]. The spiral balance spring revolutionized the accuracy of watches, enabling them to keep time to within a minute a day. This advance sparked an almost immediate rise in the market for watches, which were now no longer typically worn on a chain around the neck but were carried in a pocket. It also increased the demand for portable sundials by which watches could be set to time.

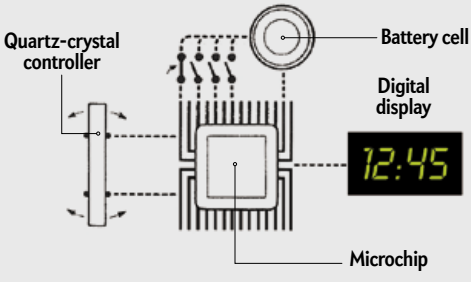
At about the same time, Huygens heard of an important English invention. The anchor escapement, unlike the verge escapement he had been using in his pendulum clocks, allowed the pendulum to swing in such a small arc that maintaining a cycloidal pathway became unnecessary. Moreover, this escapement made practical the use of a long, seconds-beating pendulum and thus led to the development of a new case design. The longcase clock, commonly

PRECISION TIMEKEEPERS

Two Modern Clocks

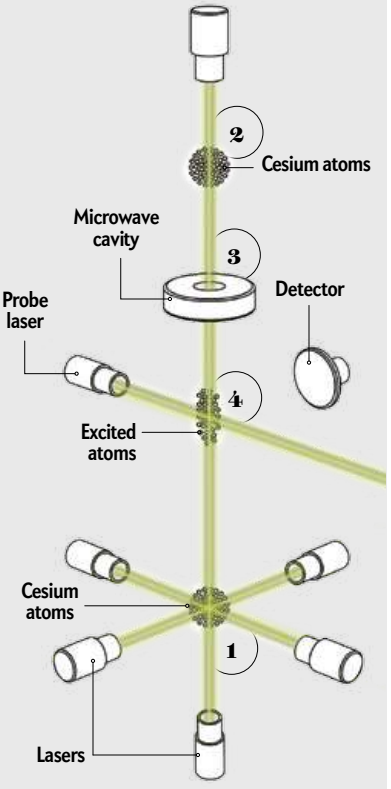
Quartz Movement

By the end of the 1960s watchmakers had taken a step away from the traditional oscillating balance wheel with the development of an electronic transistor-based oscillator comprising a tiny tuning fork whose vibrations were converted into the movement of the hands. With the simultaneous rise of cheap, low-power integrated circuits and light-emitting diodes (LEDs), the search for a more accurate timing element was on. Watchmakers soon adopted the quartz-crystal resonator from radio transmitters. Quartz crystals are piezoelectric: they vibrate when subjected to a changing electric voltage, and vice versa. When driven by a voltage at its harmonic frequency, the crystal oscillates resonantly, ringing like a bell. The output of the oscillator is then converted to pulses suitable for the watch's digital circuits, which operate an LED display or electrically actuated hands.



Cesium Fountain (Atomic) Clock

Cesium fountain clocks derive their timing reference from the frequency of an electron spin-flip transition that occurs in a cesium 133 atom when probed by tuned microwaves. In a vacuum chamber, six lasers slow the movements of gaseous cesium atoms, forming a small cloud ❶. A change in the operating frequency of the upper and lower lasers launches the atomic cloud, fountainlike ❷, up through a magnetically shielded microwave cavity ❸. As gravity pulls the cloud back down through the cavity, the electrons in the atoms interact with the microwaves for a second time. The microwaves flip the spins of the electrons, changing their quantum-mechanical energy states. After the cloud falls farther, a probe laser causes the cesium to fluoresce, revealing whether its electrons have flipped their spins, a reaction that is monitored by a detector ❹. The detector's output signal is then used to make the slight correction needed to tune the microwave emitter to a precise resonant frequency that can serve as the time beat for a clock.



DAVID PENNEY (*quartz*); ALAN DANIELS (*cesium*)

© 2018 Scientific American

known since 1876 as the grandfather clock (after a song by American Henry Clay Work), began to emerge as one of the most popular English styles. Longcase clocks with anchor escapements and long pendulums can keep time to within a few seconds a week. The celebrated English clockmaker Thomas Tompion—and, subsequently, his successor, George Graham—later modified the anchor escapement to operate without recoil. This enhanced design, called the deadbeat escapement, became the most widespread type used in precision timekeeping for the next 150 years.

SOLVING THE LONGITUDE PROBLEM

WHEN THE ROYAL OBSERVATORY, Greenwich, was founded in England in 1675, part of its charter was to find "the so-much-desired longitude of places." The first Astronomer Royal, John Flamsteed, used clocks fitted with anchor and deadbeat escapements to time the exact moments that stars crossed the celestial meridian, an imaginary line that connects the poles of the celestial sphere and defines the due-south point in the night sky. This allowed him to gather more accurate information on star positions than had hitherto been possible by making angular measurements with sextants or quadrants alone.

Although navigators could find their latitude (their position north or south of the equator) at sea by measuring the altitude of the sun or the polestar above the horizon, the heavens did not provide such a straightforward solution for finding longitude. Storms and currents often confounded attempts to keep track of distance and direction traveled across oceans. The resulting navigational errors cost seafaring nations dearly, not only in prolonged voyages but also in loss of lives, ships and cargo.

The severity of this predicament was brought home to the British government in 1707, when an admiral of the fleet and more than 1,600 sailors perished in the wrecks of four Royal Navy ships off the coast of the Scilly Isles. Thus, in 1714, through an act of Parliament, Britain offered substantial prizes for practical solutions to finding longitude at sea. The largest prize, £20,000 (roughly 200 times the annual wage of a skilled engineer of the time), would be given to the inventor of an instrument that could determine a ship's longitude to within half a degree, or 30 nautical miles, when reckoned at the end of a voyage to a port in the West Indies, whose longitude could be accurately ascertained using proved land-based methods.

The great reward attracted a deluge of harebrained schemes. Hence, the Board of Longitude, the committee appointed to review promising ideas, held no meetings for more than 20 years. Two approaches, however, had long been known to be theoretically sound. The first, called the lunar-distance method, involved precise observations of the moon's position in relation to the stars to determine the time at a reference point from which longitude could be measured; the other required a very accurate clock to make the same determination. Because the earth rotates every 24 hours, or 15 degrees in an hour, a two-hour time difference represents a 30-degree difference in longitude. The seemingly overwhelming obstacles to keeping accurate time at sea—among them the often violent motions of ships, extreme changes in temperature and variations in gravity at different latitudes—led British physicist Isaac Newton and his followers to believe that the lunar-distance method, though problematic, was the only viable solution.

Newton was wrong, however. In 1737 the board finally met for the first time to discuss the work of a most unlikely candidate, a Yorkshire carpenter named John Harrison. Harrison's large and rather cumbersome longitude timekeeper had been used on a voyage to Lisbon and on the return trip had proved its worth by correcting the navigator's dead reckoning of the ship's longitude by 68 miles. Its maker, however, was dissatisfied. Instead of asking the board for a West Indies trial, he requested and received financial support to construct an improved machine.

After two years of work, still displeased with his second effort, Harrison embarked on a third, laboring on it for 19 years. But by the time it was ready for testing, he realized that his fourth marine timekeeper, a five-inch-diameter watch he had been developing simultaneously, was better. On a voyage to Jamaica in 1761, Harrison's oversize watch performed well enough to win the prize, but the board refused to give him his due without further proof. A second sea trial in 1764 confirmed his success. Harrison was reluctantly granted £10,000. Only when King George III intervened in 1773 did he receive the remaining prize money. Harrison's breakthrough inspired further developments. By 1790 the marine chronometer was so refined that its fundamental design never needed to be changed.

MASS-PRODUCED TIMEPIECES

AT THE TURN of the 19th century, clocks and watches were relatively accurate, but they remained expensive. Recognizing the potential market for a low-cost timekeeper, two investors in Waterbury, Conn., took action. In 1807 they gave Eli Terry, a clockmaker in nearby Plymouth, a three-year contract to manufacture 4,000 longcase clock movements from wood. A substantial down payment made it possible for Terry to devote the first year to fabricating machinery for mass production. By manufacturing interchangeable parts, he completed the work within the terms of the contract.

A few years later Terry designed a wooden-movement shelf clock using the same volume-production techniques. Unlike the longcase design, which required the buyer to purchase a case separately, Terry's shelf clock was completely self-contained. The customer needed only to place it on a level shelf and wind it up. For the relatively modest sum of $15, many average people could now afford a clock. This achievement led to the establishment of what was to become the renowned Connecticut clockmaking industry.

Before the expansion of railroads in the 19th century, towns in the U.S. and Europe used the sun to determine local time. For example, because noon occurs in Boston about three minutes before it does in Worcester, Mass., Boston's clocks were set about three minutes ahead of those in Worcester. The expanding railroad network, however, needed a uniform time standard for all the stations along the line. Astronomical observatories began to distribute the precise time to the railroad companies by telegraph. The first public time service, introduced in 1851, was based on clock beats wired from the Harvard College Observatory in Cambridge, Mass. The Royal Observatory introduced its time service the next year, creating a single standard time for Britain.

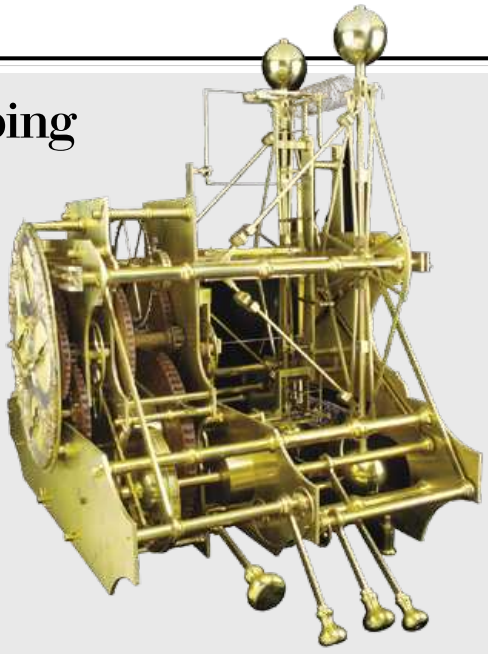
The U.S. established four time zones in 1883. By the next year the governments of all nations had recognized the benefits of a worldwide standard of time for navigation and trade. At the 1884 International Meridian Conference in Washington, D.C., the globe was divided into 24 time zones. Delegates chose the Royal Observatory as the prime meridian (zero degrees lon-

© 2018 Scientific American

Clocks That Revolutionized Timekeeping

Hemispherical Sundial (Italy, circa A.D. 100)
The shadow's direction on this ingenious Roman sundial shows the time of day (just after the eighth hour of daylight), while its length indicates the time of year (summer solstice).

Pendulum Clock (the Netherlands, 1657)
Dutch scientist Christiaan Huygens granted Salomon Coster the right to make the first pendulum clocks of his design, revolutionizing the European clockmaking industry.



H1 Sea Clock (England, 1735)
John Harrison finally solved the centuries-old problem of finding longitude at sea in 1759 with his fourth marine timekeeper. This is an exact replica of his first sea clock, called H1.

gitude, the line from which all other longitudes are measured) in part because two thirds of the world's shipping already used Greenwich time for navigation.

WATCHES FOR THE MASSES

MANY CLOCKMAKERS of this era realized that the market for watches would far exceed that for clocks if production costs could be reduced. The problem of mass-fabricating interchangeable parts for watches, however, was considerably more complicated because the precision demanded in making the necessary miniaturized components was so much greater. Although improvements in quantity manufacture had been instituted in Europe since the late 18th century, European watchmakers' fears of saturating the market and threatening their workers' jobs by abandoning traditional practices stifled most thoughts of introducing machinery for the production of interchangeable watch parts.

Disturbed that American watchmakers seemed unable to compete with their counterparts in Europe, which controlled the market in the late 1840s, a watchmaker in Maine named Aaron L. Dennison met with Edward Howard, who had established a successful clock- and scale-making business in Roxbury, Mass., to discuss mass-production methods for watches. Howard and his partner gave Dennison space to experiment and develop machinery for the project. By the fall of 1852, 20 watches had been completed under Dennison's supervision. His workmen finished 100 watches by the following spring, and 1,000 more were produced a year later. By that time the manufacturing facilities in Roxbury were proving too small, so the newly named Boston Watch Company moved to Waltham, Mass., where by the end of 1854 it was assembling 36 watches a week.

The American Waltham Watch Company, as it eventually became known, benefited greatly from a huge demand for watches during the Civil War, when Union Army forces used them to synchronize operations. Improvements in fabrication techniques further boosted output and reduced prices significantly. Meanwhile other U.S. companies formed in the hope of capturing part of the burgeoning trade. The Swiss, who had previously dominated the industry, grew concerned when their exports plummeted in the 1870s. The investigator they sent to Massachusetts discovered that not only was productivity higher at the Waltham factory but production costs were less. Even some of the lower-grade American watches could be expected to keep reasonably good time. The watch was at last a commodity accessible to the masses.

Because women had worn bracelet watches in the 19th century, wristwatches were long considered feminine accoutrements. During World War I, however, the pocket watch was modified so that it could be strapped to the wrist, where it could be viewed more readily on the battlefield. With the help of a substantial marketing campaign, the masculine fashion for wristwatches caught on after the war. Self-winding mechanical wristwatches made their appearance during the 1920s.

HIGH-PRECISION CLOCKS

AT THE END of the 19th century, Sigmund Riefler, based in Munich, developed a radical new design of regulator—a highly accurate timekeeper that served as a standard for controlling others. Housed in a partial vacuum to minimize the effects of barometric pressure and equipped with a pendulum largely unaffected by temperature variations, Riefler's regulators attained an accuracy of a tenth of a second a

© 2018 Scientific American

Shelf Clock (U.S., circa 1816)
Eli Terry designed a shelf clock with interchangeable parts, giving birth to the Connecticut clockmaking industry, and granted patent rights to Seth Thomas, who made this example, one of the earliest mass-produced clocks.

Glass-Tank Regulator (Germany, 1895)
Sigmund Riefler made a significant improvement in timekeeping when he designed a clock that operated in a partial vacuum. Combined with a pendulum made from a metal alloy with very low thermal expansion, this Riefler regulator could keep time to a 10th of a second per day.

STEPHEN PITKIN

day and were thus adopted by nearly every astronomical observatory.

Further progress came several decades later, when English railroad engineer William H. Shortt designed a so-called free pendulum clock that reputedly kept time to within about a second a year. Shortt's system incorporated two pendulum clocks, one a "master" (housed in an evacuated tank) and the other a "slave" (which contained the time dials). Every 30 seconds the slave clock gave an electromagnetic impulse to, and was in turn regulated by, the master clock pendulum, which was thus nearly free from mechanical disturbances.

Although Shortt clocks began to displace Rieflers as observatory regulators during the 1920s, their superiority was short-lived. In 1928 Warren A. Marrison, an engineer at Bell Laboratories, then in New York City, discovered an extremely uniform and reliable frequency source that was as revolutionary for timekeeping as the pendulum had been 272 years earlier. Developed originally for use in radio broadcasting, the quartz crystal vibrates at a highly regular rate when excited by an electric current [*see left illustration in box on page 54*]. The first quartz clocks installed at the Royal Observatory in 1939 varied by only two thousandths of a second a day. By the end of World War II, this accuracy had improved to the equivalent of a second every 30 years.

Quartz-crystal technology did not remain the premier frequency standard for long either, however. By 1948 Harold Lyons and his associates at the National Bureau of Standards in Washington, D.C., had based the first atomic clock on a far more precise and stable source of timekeeping: an atom's natural resonant frequency, the periodic oscillation between two of its energy states [*see right illustration in box on page 54*]. Subsequent experiments in both the U.S. and England in the 1950s led to the development of the cesium-beam atomic clock. Today the averaged times of cesium clocks in various parts of the world provide the standard frequency for Coordinated Universal Time, which has an accuracy of better than one nanosecond a day.

Up to the mid-20th century, the sidereal day, the period of the earth's rotation on its axis in relation to the stars, was used to determine standard time. This practice had been retained even though it had been suspected since the late 18th century that our planet's axial rotation was not entirely constant. The rise of cesium clocks capable of measuring discrepancies in the earth's spin, however, meant that a change was necessary. A new definition of the second, based on the resonant frequency of the cesium atom, was adopted as the new standard unit of time in 1967.

The precise measurement of time is of such fundamental importance to science and technology that the search for ever greater accuracy continues. The performance of atomic clocks had been improving by a factor of at least 10 per decade for about 50 years. But over the past decade improvements in atomic clock accuracy have dramatically accelerated. Recent advances in laser science—particularly the Nobel Prize–winning development of femtosecond laser frequency combs—and atomic physics have enabled the development of many new types of optical atomic clocks, some based on transitions in single ions in electromagnetic traps and some based on collections of cold neutral atoms held in lattices formed by laser light. Several of these atomic clocks are already stable to within a few hundred femtoseconds per day and continue to rapidly improve.

At this level of performance, formerly negligible effects become important and measurable. For example, the best atomic clocks can now measure changes in gravity over the distance of a stair step, tiny magnetic fields generated by heart and brain activity, and other quantities such as temperature and acceleration. Companies are now manufacturing "chip-scale" atomic clocks the size of a quarter. In addition to keeping time with increasing accuracy, new generations of atomic clocks will be used as exquisite sensors for myriad applications and will become ever smaller and more portable.

Although our ability to measure time will surely improve in the future, nothing will change the fact that it is the one thing of which we will never have enough. SA

MORE TO EXPLORE

Greenwich Time and the Discovery of the Longitude. Derek Howse. Oxford University Press, 1980.

History of the Hour: Clocks and Modern Temporal Orders. Gerhard Dohrn-van Rossum. Translated by Thomas Dunlap. University of Chicago Press, 1996.

The Quest for Longitude: Proceedings of the Longitude Symposium, Harvard University, Cambridge, Massachusetts, November 4–6, 1993. Edited by William J. H. Andrewes. Collection of Historical Scientific Instruments, Harvard University, 1996.

Time: From Earth Rotation to Atomic Physics. Dennis D. McCarthy and P. Kenneth Seidelmann. Wiley-VCH, 2009.

scientificamerican.com/magazine/sa

© 2018 Scientific American

FROM INSTANTANEOUS TO ETERNAL

The units of time range from the infinitesimally brief to the interminably long. The descriptions given here attempt to convey a sense of this vast chronological span

ATTOSECOND

(A billionth of a billionth of a second.) The most fleeting events that scientists can clock are measured in attoseconds. Researchers have created light pulses lasting just 53 attoseconds using high-speed lasers. Although the interval seems unimaginably brief, it is an aeon compared with the Planck time—about 10^{-43} second—which is believed to be the shortest possible duration.

FEMTOSECOND

(A millionth of a billionth of a second.) An atom in a molecule typically completes a single vibration in 10 to 100 femtoseconds. Even fast chemical reactions generally take hundreds of femtoseconds to complete. The interaction of light with pigments in the retina—the process that allows vision—takes about 200 femtoseconds.

PICOSECOND

(A thousandth of a billionth of a second.) The bottom quark, a rare subatomic particle created in high-energy accelerators, lasts for one picosecond before decaying. The average lifetime of a hydrogen bond between water molecules at room temperature is three picoseconds.

NANOSECOND

(A billionth of a second.) A beam of light shining through a vacuum will travel only 30 centimeters (not quite one foot) in this time. The microprocessor inside a personal computer will take less than a nanosecond to execute a single instruction, such as adding two numbers. The K meson, another rare subatomic particle, has a lifetime of 12 nanoseconds.

MICROSECOND

(A millionth of a second.) That beam of light will now have traveled 300 meters, about the length of three football fields, but a sound wave at sea level will have propagated only one third of a millimeter. The flash of a high-speed commercial stroboscope lasts about one microsecond. It takes 24 microseconds for a stick of dynamite to explode after its fuse has burned down.

MILLISECOND

(A thousandth of a second.) The shortest exposure time in a cell-phone camera. A housefly flaps its wings once every three milliseconds; a honeybee does the same once every five milliseconds. The moon travels around Earth two milliseconds more slowly each year as its orbit gradually widens. In computer science, an interval of 10 milliseconds is known as a jiffy.

TENTH OF A SECOND

The duration of the fabled "blink of an eye." The human ear needs this much time to discriminate an echo from the original sound. Voyager 1, a spacecraft speeding out of the solar system, travels about two kilometers farther away from the sun during this time frame. A tuning fork pitched to A above middle C vibrates 44 times.

© 2018 Scientific American

SECOND

A healthy person's heartbeat lasts about this long. On average, Americans eat 350 slices of pizza during this time. Earth travels 30 kilometers around the sun, while the sun zips 274 kilometers on its trek through the galaxy. It is not quite enough time for moonlight to reach Earth (1.3 seconds). Traditionally the second was the 60th part of the 24th part of a day, but science has given it a more precise definition: it is the duration of 9,192,631,770 cycles of one type of radiation produced by a cesium 133 atom.

MINUTE

The brain of a newborn baby grows one to two milligrams in this time. A shrew's fluttering heart beats 1,000 times. The average person can speak about 150 words or read about 250 words. When Mars is closest to Earth, sunlight reflected off the Red Planet's surface reaches us in about four minutes.

HOUR

Reproducing cells generally take about this long to divide into two. One hour and 16 minutes is the average time between eruptions of the Old Faithful geyser in Yellowstone National Park. Light from Pluto arrives at Earth in five hours and 20 minutes.

DAY

For humans, this is perhaps the most natural unit of time, the duration of Earth's rotation. Currently clocked at 23 hours, 56 minutes and 4.1 seconds, our planet's rotation is constantly slowing because of gravitational drag from the moon and other influences. The human heart beats about 100,000 times in a day, and the lungs inhale about 11,000 liters of air. In the same amount of time, an infant blue whale adds another 91 kilograms (200 pounds) to its bulk.

YEAR

Earth makes one circuit around the sun and spins on its axis 365.26 times. The mean level of the oceans rises between one and 2.5 millimeters, and North America moves about three centimeters away from Europe. It takes 4.3 years for light from Proxima Centauri, the closest star, to reach Earth—approximately the same amount of time that ocean-surface currents take to circumnavigate the globe.

CENTURY

The moon recedes from Earth by another 3.8 meters. Standard compact discs are expected to degrade in this time. Baby boomers have a one-in-26 chance of living to the age of 100, but giant tortoises can live as long as 177 years. The most advanced recordable CDs may last more than 200 years.

MILLION YEARS

After traveling for a million years, a spaceship moving at the speed of light would not yet be halfway to the Andromeda galaxy (2.3 million light-years away). The most massive stars, blue supergiants that are millions of times brighter than the sun, burn out in about this much time. Because of the movement of Earth's tectonic plates, Los Angeles will creep about 40 kilometers north-northwest of its present location in a million years.

BILLION YEARS

It took approximately this long for the newly formed Earth to cool, develop oceans, give birth to single-celled life and exchange its carbon dioxide–rich early atmosphere for an oxygen-rich one. Meanwhile the sun orbited four times around the center of the galaxy. Because the universe is 12 billion to 14 billion years old, units of time beyond a billion years aren't used very often. But cosmologists believe that the universe will probably keep expanding indefinitely, until long after the last star dies (100 trillion years from now) and the last black hole evaporates (10^{100} years from now). Our future stretches ahead much further than our past trails behind.

David Labrador, freelance writer and researcher, assembled this list.

© 2018 Scientific American

Rock of ages: Stones represent planets in the Orrery, an eight-foot-tall clock prototype showing the relative positions of Earth and its five nearest neighbors. The Long Now Foundation is building a 200-foot-tall version with huge horizontal gears and stone counterweights, designed to run for 10,000 years.

© 2018 Scientific American

DAN WINTERS

ENGINEERING

Ultimate Clocks

Atomic clocks are shrinking to microchip size, heading for space—and approaching the limits of useful precision

By W. Wayt Gibbs

IN BRIEF

A renaissance under way in atomic clock building is expected to improve the precision of timekeeping by 1,000-fold. Scientists have invented atomic clocks that are even more precise than cesium clocks—so accurate, in fact, that they can no longer measure time in seconds, which are defined by the properties of cesium.

In theory, one can measure time with infinite accuracy. But gravity and motion distort time, imposing a practical limit to clocks' precision.

Atomic clocks are short-lived. Engineers are also designing a mechanical clock that could operate through the year 12,000.

© 2018 Scientific American

W. Wayt Gibbs is a contributing editor for *Scientific American*.

DOZENS OF THE TOP CLOCKMAKERS IN THE WORLD CONVENED IN NEW ORLEANS one muggy week in May 2002 to present their latest inventions. There was not a mechanic among them; these were scientists, and their conversations buzzed with talk of spectrums and quantum levels, not gears and escapements. Today those who would build a more accurate clock must advance into the frontiers of physics and engineering in several directions at once. They are cobbling lasers that spit out pulses a quadrillionth of a second long together with chambers that chill atoms to a few millionths of a degree above absolute zero. They are snaring individual ions in tar pits of light and magnetism and manipulating the spin of electrons in their orbits.

Thanks to major technical advances, the art of ultraprecise timekeeping is progressing with a speed not seen for 30 years or more. These days a good cesium beam clock, of the kind Microsemi sells for about $50,000, will tick off seconds true to about a microsecond a month, its frequency accurate to five parts in 10^{13}. The primary time standard for the U.S., a cesium fountain clock installed in 2014 by the National Institute of Standards and Technology (NIST) at its Boulder, Colo., laboratory, is good to three parts in 10^{16} (usually written simply as 10^{-16}). That is 2,000 times the accuracy of NIST's best clock in 1975. Successful prototypes of new clock designs—devices that extract time from aluminum or mercury ions instead of cesium—have recently attained accuracy in the 10^{-18} range, a 100-fold improvement in a decade.

Accuracy may not be quite the right word. The second was defined in 1967 by international fiat to be "the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom." Leave aside for the moment what that means: the point is that to measure a second, you have to look at cesium. The best clocks now don't—so, strictly speaking, they don't measure seconds. That is one predicament the clockmakers face.

Further down the road lies a more fundamental limitation: as Albert Einstein theorized and experiment has confirmed, time is not absolute. The rate of any clock slows down when gravity gets stronger or when the clock moves quickly relative to its observer—even a single photon emitted as an electron reorients its magnetic poles or jumps from one orbit to another. By putting ultraprecise clocks on the International Space Station, scientists hope to put relativity theory through its toughest tests yet. But now that clocks have achieved a precision of 10^{-18}—proportions

PORTABLE PRECISION

Atomic Micro Clocks

"For less than $100, I could build a 10-watt jammer, drop it in New York, and block all GPS signals in the city," says Donald Sullivan, formerly of the National Institute of Standards and Technology. Navigation of all kinds depends on the Global Positioning System; smaller atomic clocks could make it more reliable. Shrunk to microchip size, they could be put into GPS receivers. The extra precision would allow the system to work on a much smaller frequency range, frustrating would-be jammers.

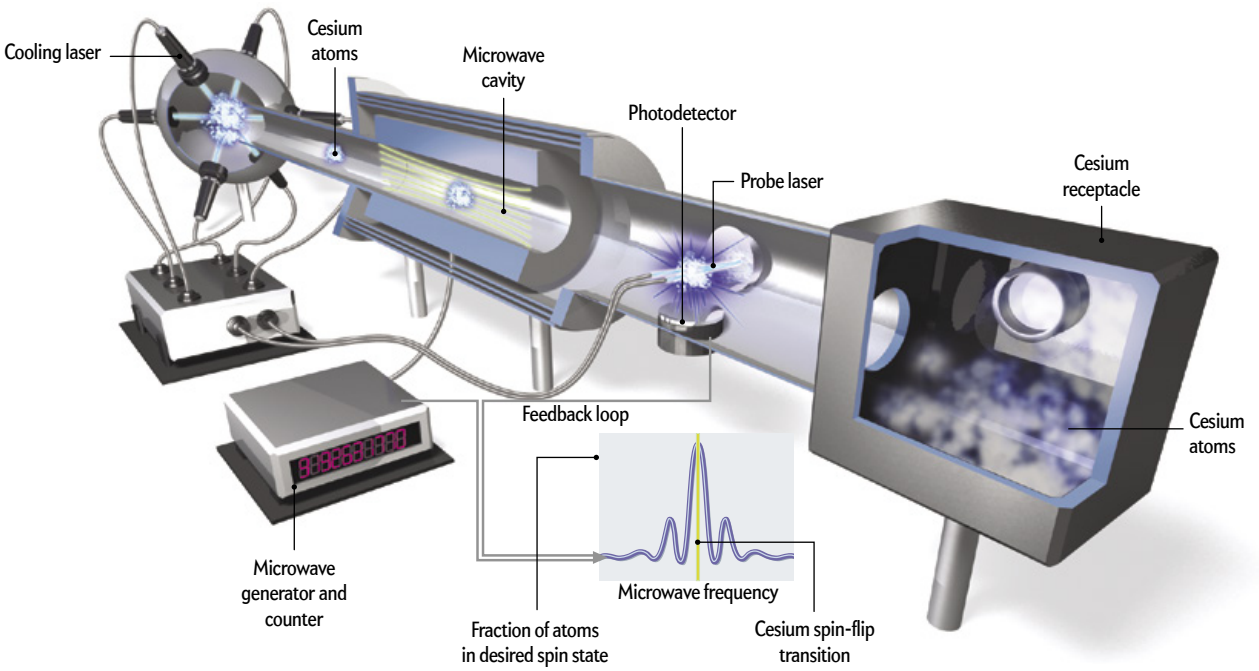
"DARPA [the Defense Advanced Research Projects Agency] has a $20-million program to develop an atomic clock on a chip for encrypted communications and GPS receivers," Sullivan reports. NIST scientists built a prototype in 2005 that is the size of a grain of rice and accurate to five parts in 10^{11}. In 2010 Symmetricom (now Microsemi) released a commercial atomic clock on a chip that is just the size of a dinner mint (*shown here*) yet matches that accuracy. —*W.W.G.*

MICROSEMI CORPORATION

© 2018 Scientific American

The Final Frontier?

PHARAO, an atomic clock (*shown here*) built by the French National Space Studies Center and other laboratories as part of a mission called ACES, has been tested on zero-gravity airplane flights. It is scheduled to launch in 2018 and be installed on the International Space Station. PHARAO aims to keep time more accurately than any cesium clock on earth. Cesium atoms, supercooled into gaseous balls by lasers, are launched through a microwave cavity, which alters the spin of their electrons. A probe laser zaps the atoms again to reveal how many were put into the desired state. A feedback loop adjusts the microwave frequency until it locks onto the natural resonance of the cesium atom spin-flip transition, which steadies the clock's "ticker." Electronics can then count 9,192,631,770 microwave cycles—exactly one second, by international consensus.



© CNES/HERVÉ PIRAUD (*PHARAO*)

that correspond to a deviation of less than half a second over the age of the universe—the effects of relativity are starting to test the scientists. No technology exists that can synchronize clocks around the world with such exactness.

INVENTING ACCURACY

SO WHY BOTHER to improve atomic clocks? The duration of the second can already be measured to 14 decimal places, a precision 1,000 times that of any other fundamental unit. One reason to do better is that the second is increasingly *the* fundamental unit. Three of the six other basic units—the meter, lumen and ampere—are now defined in terms of the second. The kilogram and the mole may be next. "It is just a matter of time before [the kilogram] is redefined," says Richard L. Steiner of NIST. Using the famous $E = mc^2$ equation, scientists could set the unit of mass to an equivalent amount of energy, such as a collection of photons whose frequencies sum to a certain number. By improving clocks, scientists can improve measurements of much more than time.

More stable and portable clock designs could also be a big boon to navigation, enhancing the accuracy and reliability of the Global Positioning System and of Galileo, a competing system in Europe that recently began initial operations. Better clocks would help NASA track its satellites, enable utilities and communications firms to trace faults in their networks, and enhance geologists' ability to pinpoint earthquakes and nuclear bomb tests. Astronomers could use them to connect telescopes in ways that dramatically sharpen their images. Inexpensive, microchip-size atomic clocks [*see box on opposite page*] are likely to have myriad uses not yet imagined.

To understand why timekeeping has suddenly lurched into

Illustrations by Bryan Christie Design

© 2018 Scientific American

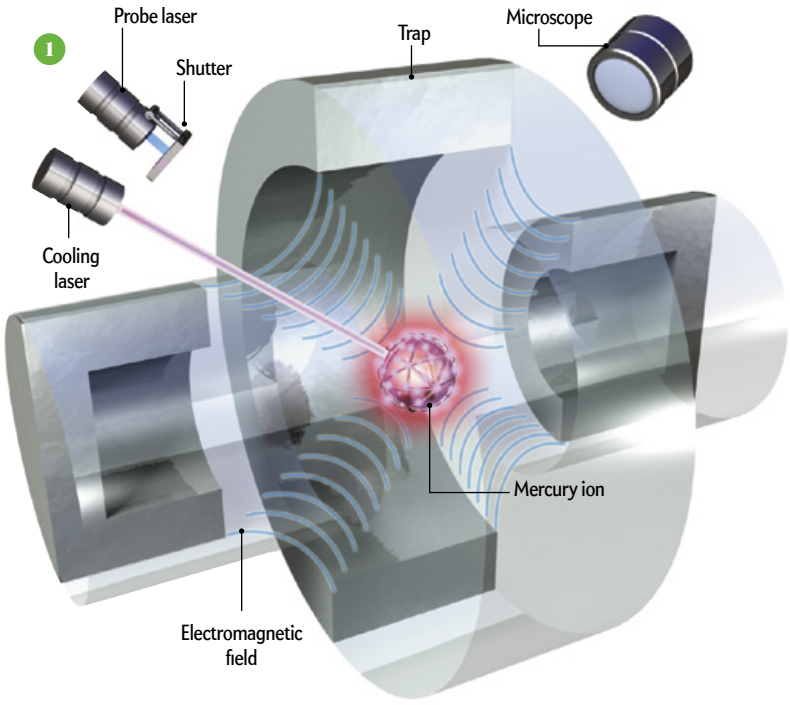
Extracting Time from an Atom

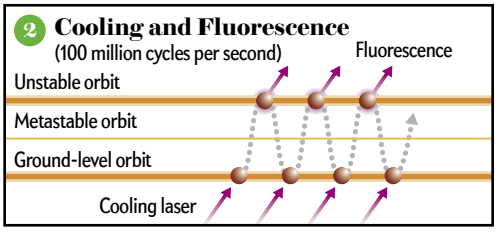
Every clock has at least two basic components, an oscillator and a counter. An atomic clock is so accurate because it includes a third element: a feedback system that periodically checks an atomic reference to keep the oscillator ticking with nearly perfect regularity.

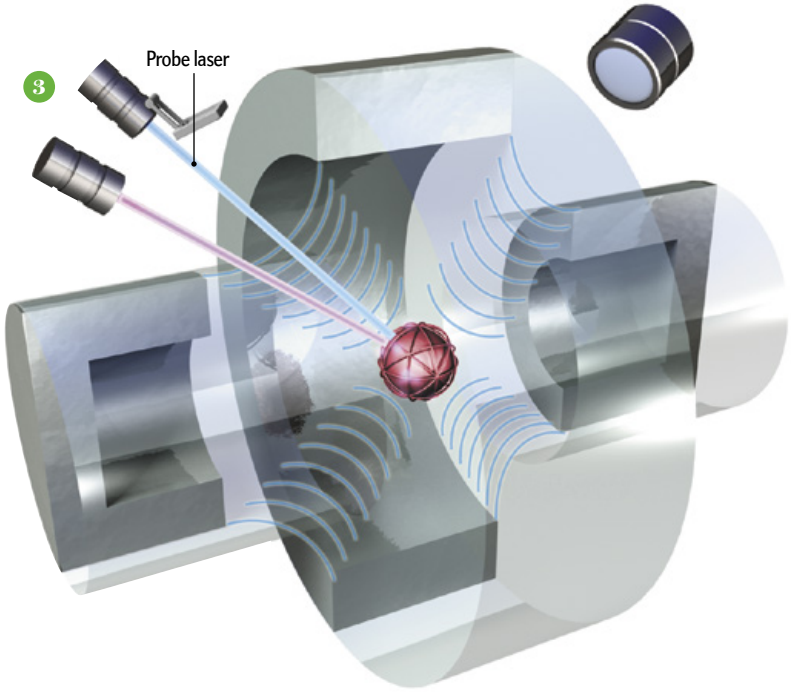
In a state-of-the-art optical ion clock, an ultraviolet probe laser serves as the oscillator. Pulses of infrared laser light yield a counter. One electron orbiting a single, nearly motionless mercury atom functions as the ultimate reference.

Trapped and Zapped

The atom, boiled off a piece of mercury in an oven, is ionized when a current strips away one of its electrons, leaving it with a positive charge. An electromagnetic field then confines the ion to the center of a ring-shaped trap (*1*). The beam of a so-called cooling laser (*purple*) causes the ion's outermost electron to jump millions of times a second to a higher, unstable orbit, fluorescing each time it falls back to the ground level (*2*). The fluorescence has two functions: it cools the atom to nearly absolute zero, and it allows scientists to verify (through a microscope) that the clock is still running. Once the atom is cool, stable and glowing, it is ready to serve as the clock's reference.

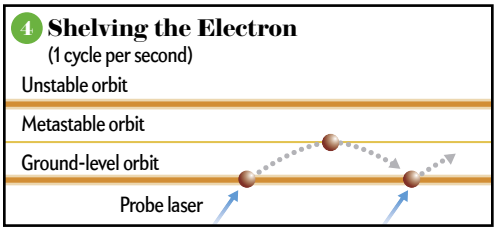






Probed and Shelved

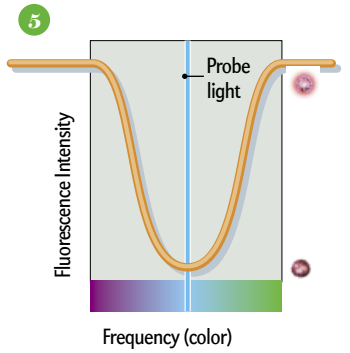
The closest thing to a ticker in an ion clock is the probe laser (*blue*). The color of the photons streaming from the laser reflects the frequency of their oscillation. To check that their frequency has not slowed or quickened, the laser periodically shines on the mercury atom (*3*). Scientists tune the color of the probe light to the precise frequency that knocks the ion's outer electron into a metastable orbit, thus "shelving" the electron for up to half a second (*4*). When the laser is tuned to this special frequency, the electron stops fluorescing, and the ion goes dark. If the laser oscillator drifts, the ion blinks back on.

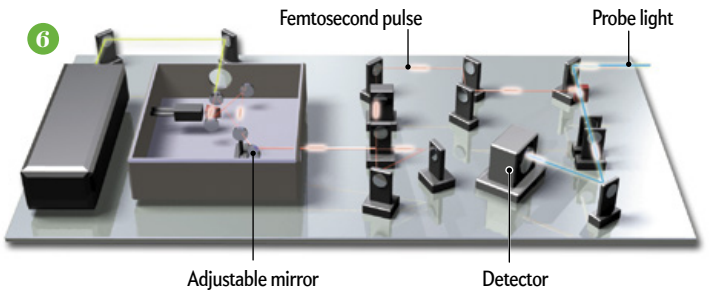


© 2018 Scientific American

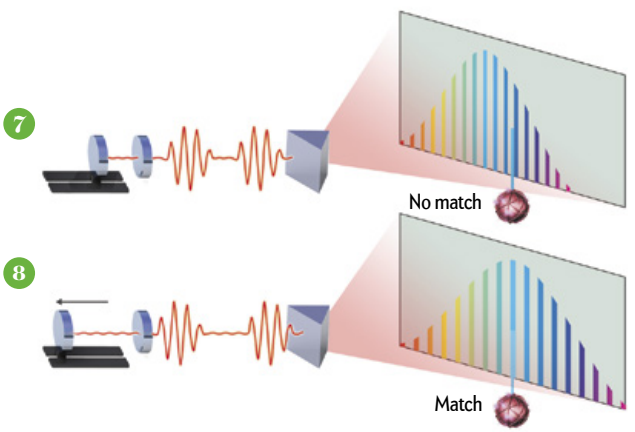
Matched and Metered

A feedback system adjusts the laser color until the fluorescence is at a minimum (5). The probe light, now rock steady, is next passed via optical fiber to a counter. The probe light oscillates about a quadrillion times a second, far too fast to count directly. A third laser acts like a reducing gear to translate the time signal from a quadrillion cycles a second to about a billion cycles a second. This third laser emits infrared pulses just a few femtoseconds long, with stretches of darkness between them (6).





The trick is to lock its pulse rate in perfect synchronicity with the frequency of the probe light. To do this, the clockwork exploits a curious fact: when passed through a prism, each ultrashort pulse splits into a rainbow of colors spaced at regular frequency intervals, like the teeth on a gear (7 and 8). By moving an adjustable mirror, scientists alter the delay between pulses, thereby stretching or compressing the range of frequencies carried by each pulse. This allows them to position the "gear" so that one of its teeth matches the color (and thus the frequency) of the probe light—which means that it is also locked to the hardwired behavior of the mercury ion. An electronic detector then counts the synchronized pulses as they go by, a billion a second, ticking off the passage of time.



high gear, it helps to know a little about how atomic clocks work. In principle, an atomic clock is just like any other timepiece, with an oscillator that "ticks" in a regular way and a counter that converts the ticks to seconds. The ticker in a cesium clock is not mechanical (like a pendulum) or electromechanical (like a quartz crystal). It is quantum-mechanical: a photon of light is absorbed by the cesium atom's outermost electron, causing the electron to flip its magnetic field (and its associated spin) upside down.

Unlike pendulums and crystals, all cesium atoms are identical. And every one will flip the spin of its outer electron when hit with microwaves at the frequency of exactly 9,192,631,770 cycles per second. To measure seconds, the clock locks its microwave generator onto the sweet spot in the spectrum where the most cesium atoms react. Then it starts counting cycles.

Of course, nothing in quantum physics is really that simple. Complicating things, as usual, is the Heisenberg indeterminacy principle, which puts strict limits on how precisely one can measure the frequency of a single photon. The best clocks now scan a one-hertz-wide sweet spot to find its exact center, plus or minus one millihertz, in every single measurement—despite the Heisenberg limits. "The reason we can do it is that we look at more than a million atoms each time," explained Kurt Gibble, a physicist at Pennsylvania State University, in New Orleans. "Because it isn't really just one measurement, it doesn't violate the laws of quantum mechanics."

But that solution creates other problems. At room temperature, cesium is a soft, silvery metal. It would melt in your palm to a golden puddle—although you wouldn't want to touch it, because it reacts violently with water. Inside a cesium beam clock, an oven heats the metal until atoms boil off. These hot particles can zip through the microwave cavity at various speeds and angles. Some move so fast that (because of relativity) they behave as if time has slowed. To other atoms, the microwaves appear (because of Doppler shifting) to be higher or lower in frequency than they are. The atoms no longer behave identically, so the ticks grow less distinct.

Herr Doktor Heisenberg would probably have suggested slowing the atoms down, and that's what clockmakers have done. Several clocks in the world—at NIST, the U.S. Naval Observatory, and the standards institutes in Paris, Teddington, England, and Brunswick, Germany—toss supercooled balls of cesium atoms in a fountainlike arc through a microwave chamber [see "A Chronicle of Timekeeping," by William J. H. Andrewes, on page 50]. To condense the hot cesium gas into a ball, six intersecting laser beams decelerate the atoms to less than two microkelvins—almost a complete standstill. The low temperature all but eliminates relativistic and Doppler shifts, and it gives a two-meter-tall fountain clock half a second to flip the atoms' spins. Fountain clocks, introduced in 1996, rapidly knocked 90 percent off the uncertainty of international atomic time.

© 2018 Scientific American

A Clock for All Time

SAN RAFAEL, CALIF., 2002—A NIST Web site boasts that an atomic chronometer it has constructed "would neither gain nor lose one second in about 3.7 billion years." Atomic horologists often speak as if their timepieces could run continuously for thousands of centuries. Balderdash—a typical cesium clock lasts no more than 20 years. A decent wristwatch runs longer.

But at facilities in California and Seattle, a small group of futurists and engineers have begun building a mechanical clock meant to tick through 1,000 decades. The Clock of the Long Now, as its chief designer, Danny Hillis, calls it, is as much a sociological experiment as a functional chronometer. (Hillis serves on *Scientific American*'s board of advisers.)

"A clock is a symbol of continuity; one that lasts a really long time might give people a sense of perspective, help them think about the year 3000 as more than just an abstraction," Hillis says. "Our record of civilization extends back roughly 10,000 years, so that struck me as a good interval to look forward."

Hillis may seem like an unlikely leader of a movement to reverse society's preoccupation with the fast and soon. In the 1980s he designed supercomputers; in the 1990s it was theme park rides. He can spare an hour for an interview only if half of it is done on the trip to Silicon Valley for his next meeting.

Nevertheless, Hillis, with help from writer Stewart Brand, musician Brian Eno and others, is trying to craft an artifact that will not just endure but will also inspire. The clock will have to be wound once a year. "And when you first come up to it, it will only display what time it was when the last person was there," Hillis explains. "It will track the current time, but you will have to wind it—put some energy into it—to get it to advance to show what time it is now."

Brand and Hillis co-chair a foundation (www.longnow.org) that has been given use of a mountain in Texas on land owned by Jeff Bezos, inside which they are constructing the monument-size clock. Through a slit in the cavern ceiling, rays of the noon sun will focus on a bimetallic strip, triggering a weight to resynchronize the clock in case its time has drifted.

Although this all may sound quite spiritual, "we don't want to create a religion," Brand avers as he stands next to a mock-up of the second prototype. This version is twice the size of the first, on exhibit at the Science Museum in London. In place of a circular dial, however, the clock is now crowned with a large orrery indicating planetary positions.

Below the "face" sits a stack of seven metal rings, each about 75 centimeters in diameter and fringed with levers. Vertical pins stuck into the rings engage the levers as the rings rotate, working as a mechanical binary computer to count the hours and compute the date. Because the clockwork is mechanical and is open to inspection, "you can figure out how to restart it if it hasn't been on in 100-odd years," Hillis says. But whether his idea gathers enough currency to get a 10,000-year clock started in the first place, only time will tell. —*W.W.G.*

10,000-year clock under construction by the Long Now Foundation will be strictly mechanical. Like this first prototype of the clock, the final, monument-size version will probably use a torsional pendulum to count minutes but will display only the current year, century and millennium.

COURTESY OF ROLFE HORN *Long Now Foundation (http://www.longnow.org)*

TIME IN SPACE

IT TAKES TIME to make a good second, and the fountain clocks still rush the job. "We would have to quadruple the height of the tower to double the observation time," says Donald Sullivan, former chief of the time and frequency division at NIST. Instead of punching a hole through the ceiling of his lab, Sullivan led one of three projects to put fountainlike clocks on the International Space Station. "In space, we can launch a ball of atoms at 15 centimeters per second through a 74-centimeter cavity. So we have five to 10 seconds to observe them," he explains. The $25-million Primary Atomic Reference Clock in Space (PARCS) project on which he worked was designed to turn out seconds good to five parts in 10^{17}.

PARCS was canceled in 2004, when NASA shifted funding from the space station to programs to send astronauts to the moon and, eventually, Mars. But a device from the European Space Agency called ACES (Atomic Clock Ensemble in Space) is scheduled to launch in 2018 and aims to measure with 99.99997 percent accuracy how much the microgravity of low Earth orbit slows time compared with measurements made on the ground.

Meanwhile efforts to make a third space-faring clock, called RACE (Rubidium Atomic Clock Experiment), helped to refine a newer approach that replaces the cesium so familiar to clockmakers with a different alkali element. "In the best cesium fountains the largest source of error are so-called cold collisions," explained Gibble—who directed the RACE project, which died along with PARCS in 2004—in New Orleans. At temperatures near absolute zero, quantum physics takes over and atoms start to behave like waves. "They appear hundreds of times bigger than normal, so they collide much more often. At a microkelvin, cesium has nearly the maximum possible cross section," he continued. "But the effective size for rubidium atoms is 50 times smaller." That enables rubidium clocks to reach 10^{-17}, one fifth the uncertainty of ACES.

Rubidium clocks offer another advantage: the opportunity to look for fluctuations in the fine-structure constant, alpha (α). Alpha determines the strength of electromagnetic interactions in atoms and molecules. It is very nearly 1/137, a unitless num-

© 2018 Scientific American

ber that falls out of the Standard Model of physics, with no apparent reason for the value it has. Yet it is an important number—change α very much, and the universe could not support life as we know it [see "Inconstant Constants," by John D. Barrow and John K. Webb, on page 70].

In the Standard Model, the fine-structure constant is immutable throughout eternity. But in some competing theories (such as certain string theories), α could waver slightly or grow as time goes by. In August 2001 a group of astronomers reported preliminary evidence that α may have increased by one part in 10,000 during the past six billion years. But the evidence is equivocal, and the question is a hard one to settle. By analyzing radio signals from a distant galaxy, astronomers in February concluded that if α is fluctuating, it has varied by no more than about one part in a million over the past 2.9 billion years.

LASERS RULE

INDEED, THE DEVELOPMENT of ion clocks in recent years has made clocks based on fountains of atoms seem almost obsolete. In August 2001 Scott A. Diddams and his colleagues at NIST reported an initial trial run of something many clock builders had thought they might never live to see: an optical atomic clock based on a single mercury atom. It may seem like a natural idea to graduate from microwaves, at frequencies of gigahertz, to visible light, well into the terahertz part of the spectrum. Optical photons pack enough energy to bump electrons clear into the next orbital shell—no need to fuss with subtleties like spin. But although the ticker still works at terahertz frequencies, the counter breaks.

"Nobody knows how to count 10^{16} cycles per second," observes Eric A. Burt of NASA's Jet Propulsion Laboratory. "We needed a bridge to the microwave regime, where we do have electronic counters."

Enter the optical ruler. In 1999 Thomas Udem, Theodor W. Hänsch and others at the Max Planck Institute for Quantum Optics in Garching, Germany, figured out a way to measure optical frequencies directly, using a reference laser that pulses at a rate of one gigahertz. Each pulse of light is just a couple of dozen femtoseconds long. (A femtosecond is a very, very small amount of time. More femtoseconds elapse in each second than there have been hours since the big bang.) A laser puts out a continuous beam of only one color, but pulse that laser, and you get a mixture of colors in each flash. The spectrum of a femtosecond pulse is a bizarre thing to see: millions of sharp lines spanning the rainbow, each line spaced exactly the same distance from its neighbors—like tick marks on a ruler. "That you could make a laser that pulses a billion times a second and whose constituent frequencies are all stable to one hertz is just short of unbelievable," Gibble said in New Orleans, shaking his head.

Diddams's group at NIST has built a rudimentary optical clockwork around mercury ions, which they immobilize in an electromagnetic trap [*see box on pages 64 and 65*]. Because each atom is missing an electron, the ions carry a positive charge. They repel one another, so collisions are no longer a problem. The device is stable to better than six parts in 10^{16} over the course of a second. Over longer periods the uncertainty could approach 10^{-18}. "Mercury is not an ideal element to use," Sullivan acknowledges. "The clock transition we use in it can shift with magnetic fields, which are hard to eliminate completely. But there is a transition in indium that looks attractive."

Udem and Hänsch are one step ahead of him. They have been investigating the indium ion, and indeed it seems quite capable of carrying clocks down "into the 18's," as Gibble put it. Groups at the Federal Institute of Physics and Metrology in Brunswick, Germany, and elsewhere are experimenting with uncharged calcium atoms. Because neutral atoms can be crammed more densely into the trap than ions can, the signal soars higher over the noise. And in 2015 a NIST-led team reported the successful demonstration of an optical lattice clock based on atoms of strontium 87; the uncertainty was pegged at two parts in 10^{18}. A German research group reported in 2016 on a precision almost as good, of 3.2 parts in 10^{18}, from a system based on a single ion of ytterbium 171. The NIST group also used ytterbium for a lattice clock that achieved an accuracy of 1.6 parts in 10^{18}.

INCONSTANT TIME

BUT THERE'S THAT WORD AGAIN: accuracy. These new clocks "move away from the atomic definition of the second, which is based on the properties of cesium," Sullivan points out. For the newest and best clocks to be strictly accurate as keepers of the time to which we set our watches, that definition will have to change. Sullivan says the time committee of the International Bureau of Weights and Measures (BIPM), which decides such things, accepted his proposal to allow "secondary" definitions that state the equivalence of a cesium frequency to that of other atoms. If the full BIPM assembly approves the idea, the definition of the second will be broadened but also weakened.

Clock builders will not get around relativity so easily. Clocks accurate to one part in 10^{17}—a millisecond in three million years—will be easily thrown out of whack by two relativistic effects. First, there is time dilation: moving clocks run slow. "A frequency shift of 10^{-17} corresponds to a time dilation due to walking speed," Gibble said.

The other confounder is gravity. The stronger its pull, the slower time passes. Clocks at the top of Mount Everest pull ahead of those at sea level by about 30 microseconds a year. When NIST researchers synchronized two quantum logic clocks and then raised one of them 33 centimeters, they found that the clocks were perceptibly out of sync. Raising a clock 10 centimeters changes its rate by one part in 10^{17}. And elevation is relatively easy to measure, compared with variations in gravity caused by local geology, tides or even magma shifting kilometers underground.

Ultimately, Gibble said, "if you take our ability to split spectral lines with microwave clocks and extrapolate to optical rulers, that puts you at uncertainties of order 10^{-22}. I certainly would not claim that we are going to get there anytime soon, however." There is no particular rush: no one has the first idea how to transfer time that precisely between two clocks. And what good is a clock if you can't move it and can't check it against another? SA

MORE TO EXPLORE

Ultrashort-Pulse Lasers. John-Mark Hopkins and Wilson Sibbett in *Scientific American,* Vol. 283, No. 3, pages 72–79; September 2000.

Splitting the Second: The Story of Atomic Time. Tony Jones. Institute of Physics Publishing, 2000.

An Optical Clock Based on a Single Trapped 199Hg+ Ion. Scott A. Diddams et al. in *Science,* Vol. 293, pages 825–828; August 3, 2001.

The measurement of time: www.npl.co.uk/educate-explore/what-is-the-time

NIST Time and Frequency Division: http://tf.nist.gov

© 2018 Scientific American

John Matson is a former reporter and editor for *Scientific American* who has written extensively about astronomy and physics.

RELATIVITY

HOW TIME FLIES

Some experimental optical clocks are so precise that even a small change in elevation or velocity makes them register the passage of time differently

By John Matson

IF YOU HAVE EVER FOUND YOURSELF CURSING A NOISY UPSTAIRS NEIGHBOR, take solace in the fact that he or she is aging faster than you are.

Einstein's general theory of relativity predicts that clocks at different gravitational potentials will tick at different rates—a clock at higher elevation will tick faster than will a clock closer to Earth's center. In other words, time passes more quickly in your neighbor's upstairs apartment than it does in your apartment.

IN BRIEF

The development of ultraprecise atomic clocks has made it possible to demonstrate relativistic time effects in the laboratory.
Recent experimental atomic clocks rely on aluminum ions, which can keep time to within one second in roughly 3.7 billion years.
In the past relativistic effects could be detected only at massive scales of distance or velocity or by harnessing atomic oscillations too rapid to be accurately counted.

To complicate matters, Einstein's special theory of relativity, which preceded general relativity by a decade, predicts a similar effect for clocks in motion—a stationary clock will tick faster than a moving clock. This is the source of the famous twin paradox: following a round-trip journey on a spaceship traveling at some exceptionally high velocity, a traveler would return to Earth to find that her twin sibling is now older than she is, because time has passed more slowly on the moving ship than on Earth.

Both these so-called time dilation effects have been verified in a number of experiments throughout the decades, which have traditionally depended on large scales of distance or velocity. In one landmark test, in 1971, Joseph C. Hafele of Washington University in St. Louis and Richard E. Keating of the U.S. Naval Observatory flew cesium atomic clocks around the world on commercial jet flights, then compared the clocks with reference clocks on the ground to find that they

© 2018 Scientific American

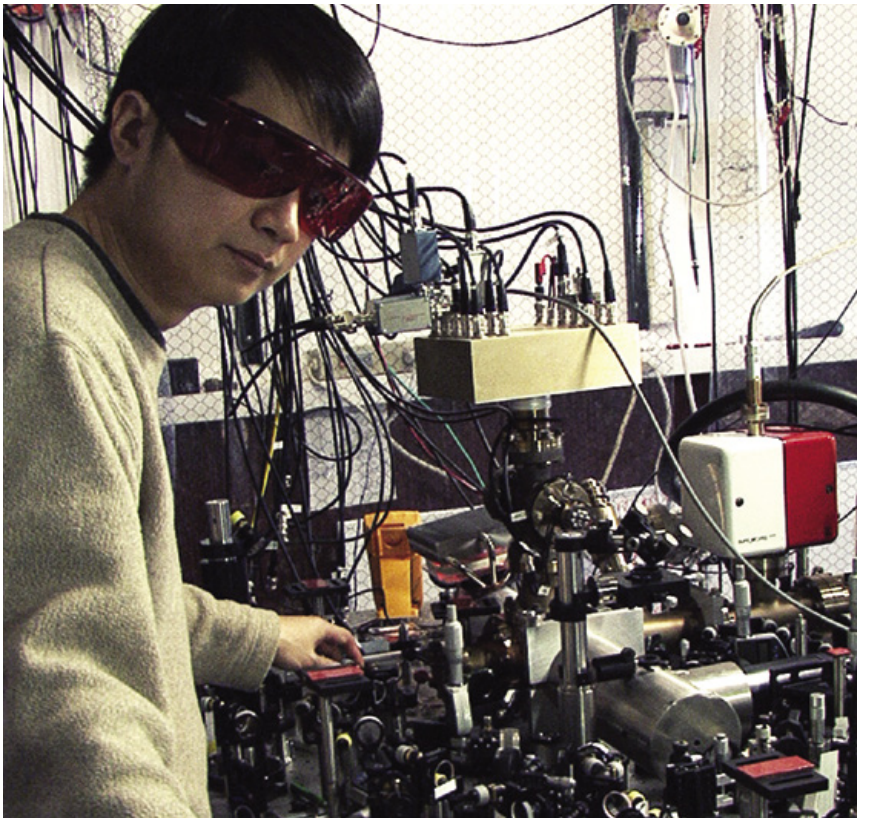
had diverged, as predicted by relativity. Yet even at the speed and altitude of jet aircraft, the effects of relativistic time dilation are tiny—in the Hafele-Keating experiment, the atomic clocks differed after their journeys by just tens to hundreds of nanoseconds.

Thanks to improved timekeeping, similar demonstrations can now take place at more mundane scales in the laboratory. In a series of experiments described in the September 24, 2010, *Science,* researchers at the National Institute of Standards and Technology (NIST) registered differences in the passage of time between two high-precision optical atomic clocks when one was elevated by just a third of a meter or when one was set in motion at speeds of less than 10 meters per second.

Again, the effects are minuscule: it would take the elevated clock hundreds of millions of years to log one more second than its counterpart, and a clock moving a few meters per second would need to run about as long to lag one second behind its stationary counterpart. But the development of optical clocks based on aluminum ions, which can keep time to within one second in roughly 3.7 billion years, allows researchers to expose those diminutive relativistic effects. "People usually think of it as negligible, but for us it is not," says lead study author James Chin-wen Chou of NIST. "We can definitely see it."

The NIST group's optical clocks use lasers to probe the quantum state of aluminum ions held in radio-frequency traps. When the laser's frequency is just right, it resonates with a transition between quantum states in the aluminum ion whose frequency is constant in time. By continually tuning the laser to drive that aluminum transition, an interaction that occurs only in a tiny window near 1.121 petahertz (1.121 quadrillion cycles per second), the laser's frequency can be stabilized to an exquisitely sensitive degree, allowing it to act as the clock's pendulum. "If we anchor the frequency of the oscillator—in our case, laser light—to the unchanging, stable optical transition in aluminum, the laser oscillation can serve as the tick of the clock," Chou explains.

To put the sensitivity of the optical clocks in perspective, Chou notes that the two timekeepers in the study differed after a height change of a mere step on a staircase—never mind the entire floor separating you from your noisy neighbor—or with just a few meters per second of motion. "If you push your daughter on a swing, it's about that speed," he says.



COURTESY OF J. BURRUS *National Institute of Standards and Technology*

It's all relative: James Chin-wen Chou with one of the aluminum-ion optical clocks at the National Institute of Standards and Technology.

In the past such relativistic experiments have involved either massive scales of distance or velocity or else oscillations so fast that their ticks cannot be reliably counted for timing purposes, notes Holger Müller, an atomic physicist at the University of California, Berkeley. "It's an enormous achievement that you can build optical clocks so good that you can now see relativity in the lab," he says.

Müller has used atom interferometry to make precision measurements of relativistic effects—measurements that rely not on counting individual oscillations but rather on tracking the interference between quantum waves representing individual atoms. (The frequencies of such waves, which oscillate tens of billions of times faster than the petahertz laser in an aluminum clock, are simply too high to monitor and count.) It is a process akin to striking two tuning forks to listen to the pulsations of their interference without actually measuring how many times each fork vibrates. In that sense, atom interferometers are pendulums without clockwork, so although they can make physical measurements with great precision, they cannot be used to keep time.

"The new work operates on familiar scales of distance and velocity, with clocks that can be used for universal timing applications," Müller says. "They see the effects of general and special relativity, and that makes relativity something you can kind of see and touch." SA

MORE TO EXPLORE

Around-the-World Atomic Clocks: Observed Relativistic Time Gains. J. C. Hafele and Richard E. Keating in *Science,* Vol. 177, pages 168–170; July 14, 1972.

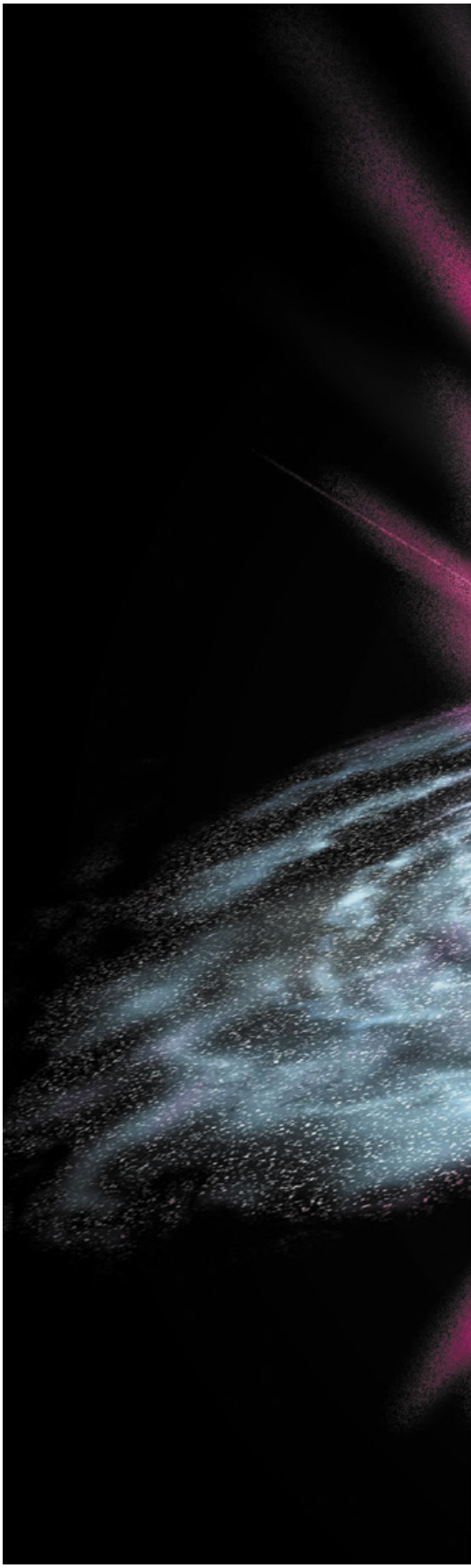
Black Holes and Time Warps: Einstein's Outrageous Legacy. Kip S. Thorne. W. W. Norton, 1994.

A Precision Measurement of the Gravitational Redshift by the Interference of Matter Waves. Holger Müller, Achim Peters and Steven Chu in *Nature,* Vol. 463, pages 926–929; February 18, 2010.

Frequency Comparison of Two High-Accuracy Al+ Optical Clocks. C. W. Chou et al. in *Physical Review Letters,* Vol. 104, No. 7; February 19, 2010. http://arxiv.org/abs/0911.4527

© 2018 Scientific American

John D. Barrow and **John K. Webb** began to work together to probe the constants of nature in 1996, when Webb spent a sabbatical with Barrow at the University of Sussex in England. Since 1999 Barrow has been professor of mathematical sciences at the University of Cambridge and a fellow of the Royal Society. Webb is a professor at the University of New South Wales in Sydney, Australia. Both are known for their efforts to explain science to the public.



ASTROPHYSICS

INCONSTANT CONSTANTS

Do the inner workings of nature change with time?

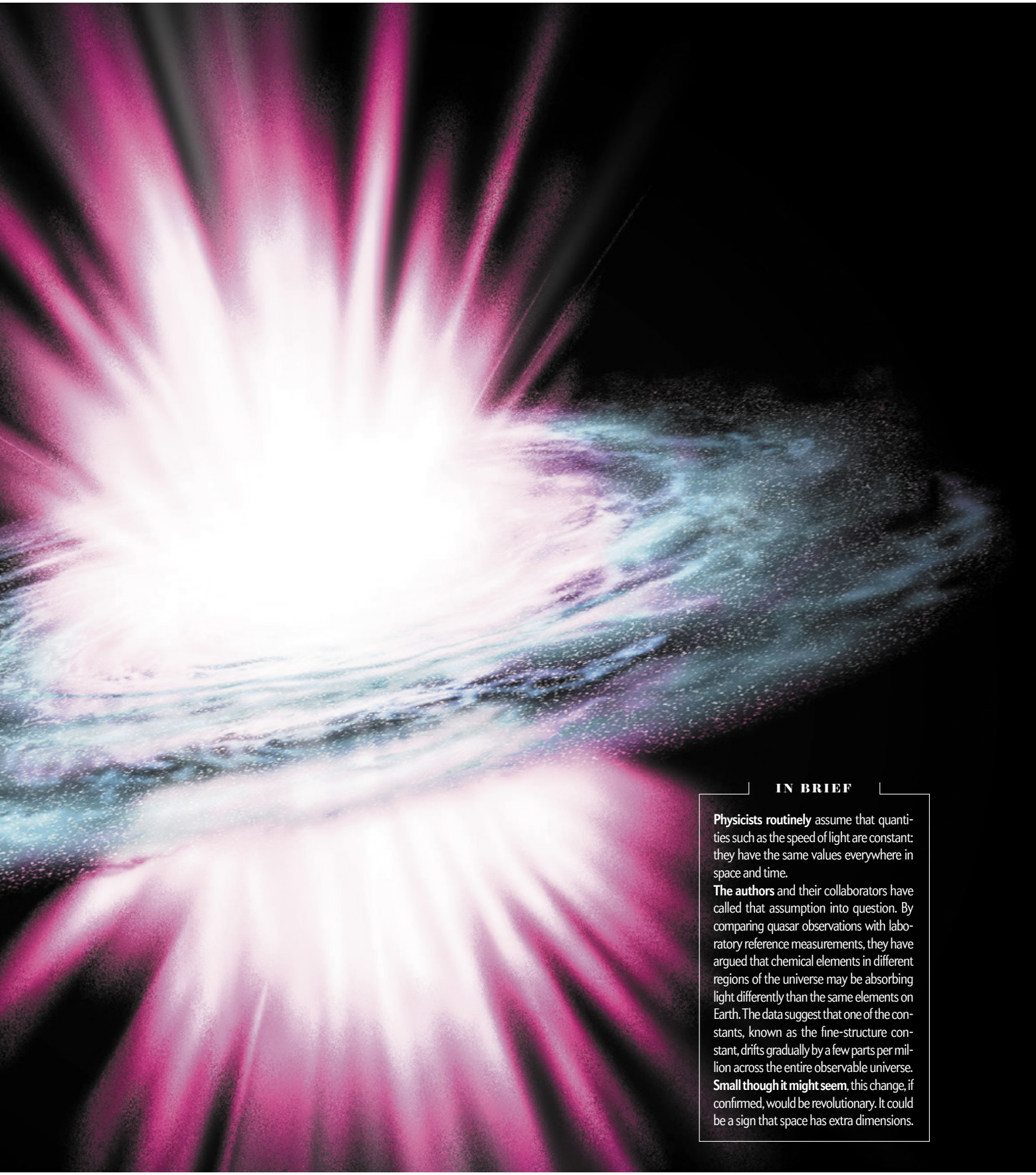
By John D. Barrow and John K. Webb

Illustration by Ron Miller

SOME THINGS NEVER CHANGE. PHYSICISTS CALL them the constants of nature. Such quantities as the velocity of light, c, Newton's constant of gravitation, G, and the mass of the electron, m_e, are assumed to be the same at all places and times in the universe. They form the scaffolding around which the theories of physics are erected, and they define the fabric of our universe. Physics has progressed by making ever more accurate measurements of their values.

And yet, remarkably, no one has ever successfully predicted or explained any of the constants. Physicists have no idea why constants take the special numerical values that they do (given the choice of units). In SI units, c is 299,792,458; G is 6.673×10^{-11}; and m_e is $9.10938188 \times 10^{-31}$—numbers that follow no discernible pattern. The only thread running through the values is that if many of them were even slightly different, complex atomic structures such as living beings would not be possible. The desire to explain

© 2018 Scientific American



IN BRIEF

Physicists routinely assume that quantities such as the speed of light are constant: they have the same values everywhere in space and time.

The authors and their collaborators have called that assumption into question. By comparing quasar observations with laboratory reference measurements, they have argued that chemical elements in different regions of the universe may be absorbing light differently than the same elements on Earth. The data suggest that one of the constants, known as the fine-structure constant, drifts gradually by a few parts per million across the entire observable universe.

Small though it might seem, this change, if confirmed, would be revolutionary. It could be a sign that space has extra dimensions.

© 2018 Scientific American

the constants has been one of the driving forces behind efforts to develop a complete unified description of nature, or "theory of everything." Physicists have hoped that such a theory would show that each of the constants of nature could have only one logically possible value. It would reveal an underlying order to the seeming arbitrariness of nature.

In recent years, however, the status of the constants has grown more muddied, not less. Researchers have found that the best candidate for a theory of everything, the variant of string theory called M-theory, is self-consistent only if the universe has more than four dimensions of space and time—as many as seven more. One implication is that the constants we observe may not, in fact, be the truly fundamental ones. Those live in the full higher-dimensional space, and we see only their three-dimensional "shadows."

Meanwhile physicists have also come to appreciate that the values of many of the constants may be the result of mere happenstance, acquired during random events and elementary particle processes early in the history of the universe. In fact, string theory allows for a vast number—10^{500}—of possible "worlds" with different self-consistent sets of laws and constants. Thus far researchers have no idea why our combination was selected. Continued study may reduce the number of logically possible worlds to just one, but we have to remain open to the unnerving possibility that our known universe is but one of many—a part of a multiverse—and that different parts of the multiverse exhibit different solutions to the theory, our observed laws of nature being merely one edition of many systems of local bylaws.

No further explanation would then be possible for many of our numerical constants other than that they constitute a rare combination that permits consciousness to evolve. Our observable universe could be one of many isolated oases surrounded by an infinity of lifeless space—a surreal place where different forces of nature hold sway and particles such as electrons or structures such as carbon atoms and DNA molecules could be impossibilities. If you tried to venture into that outside world, you would cease to be.

Thus, string theory gives with the right hand and takes with the left. It was devised in part to explain the seemingly arbitrary values of the physical constants, and the basic equations of the theory contain few arbitrary parameters. Yet so far string theory offers no explanation for the observed values of the constants.

IF WE LOOKED FAR ENOUGH OUT IN SPACE, WE MIGHT BEGIN TO SEE REGIONS WHERE THE "CONSTANTS" HAVE SETTLED INTO DIFFERENT VALUES.

A RULER YOU CAN TRUST

INDEED, THE WORD "constant" may be a misnomer. Our constants could vary both in time and in space. If the extra dimensions of space were to change in size, the "constants" in our three-dimensional world would change with them. If we looked far enough out in space, we might begin to see regions where the "constants" have settled into different values. Ever since the 1930s researchers have speculated that the constants may not be constant. String theory gives this idea a theoretical plausibility and makes it all the more important for observers to search for deviations from constancy.

Such experiments are challenging. The first problem is that the laboratory apparatus itself may be sensitive to changes in the constants. The size of all atoms could be increasing, but if the ruler you are using to measure them is getting longer, too, you would never be able to tell. Experimenters routinely assume that their reference standards—rulers, masses, clocks—are fixed, but they cannot do so when testing the constants. They must focus on constants that have no units—they are pure numbers—so their values are the same irrespective of the units system. An example is the ratio of two masses, such as the proton mass to the electron mass.

One ratio of particular interest combines the velocity of light, c, the electric charge on a single electron, e, Planck's constant, h, and the so-called vacuum permittivity, ε_0. This famous quantity, alpha (α) = $e^2/2\varepsilon_0 hc$, called the fine-structure constant, was first introduced in 1916 by Arnold Sommerfeld, a pioneer in applying the theory of quantum mechanics to electromagnetism. It quantifies the relativistic (c) and quantum (h) qualities of electromagnetic (e) interactions involving charged particles in empty space (ε_0). Measured to be equal to 1/137.03599976, or approximately 1/137, α has endowed the number 137 with a legendary status among physicists (it usually opens the combination locks on their briefcases).

If α had a different value, all sorts of vital features of the world around us would change. If the value were lower, the density of solid atomic matter would fall (in proportion to α^3), molecular bonds would break at lower temperatures (α^2), and the number of stable elements in the periodic table could increase ($1/\alpha$). If α were too big, small atomic nuclei could not exist, because the electrical repulsion of their protons would overwhelm the strong nuclear force binding them together. A value as big as 0.1 would blow carbon apart.

The nuclear reactions in stars are especially sensitive to α. For fusion to occur, a star's gravity must produce temperatures high enough to force nuclei together despite their tendency to repel one another. If α exceeded 0.1, fusion would be impossible (unless other parameters, such as the electron-to-proton mass ratio, were adjusted to compensate). A shift of just 4 percent in α would alter the energy levels in the nucleus of carbon to such an extent that the production of this element by stars would shut down.

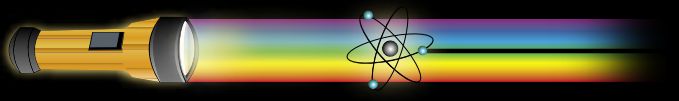
NUCLEAR PROLIFERATION

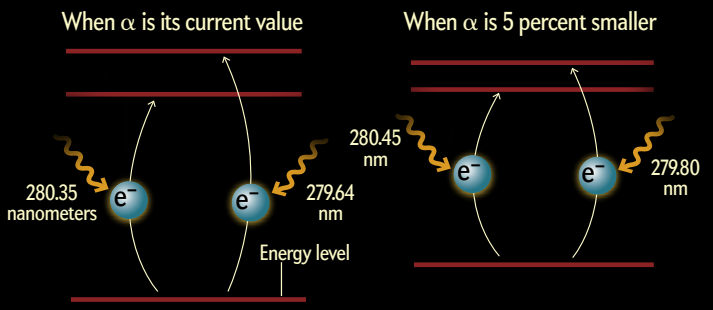
THE SECOND experimental problem, less easily solved, is that measuring changes in the constants requires high-precision equipment that remains stable long enough to register any changes. Even atomic clocks can detect drifts in the fine-structure constant only over days or, at most, years. If α changed by more than four parts in 10^{15} over a three-year period, the best clocks would see it. None have. That may sound like an impressive confirmation of constancy, but three years is

© 2018 Scientific American

Light and the Fine-Structure Constant

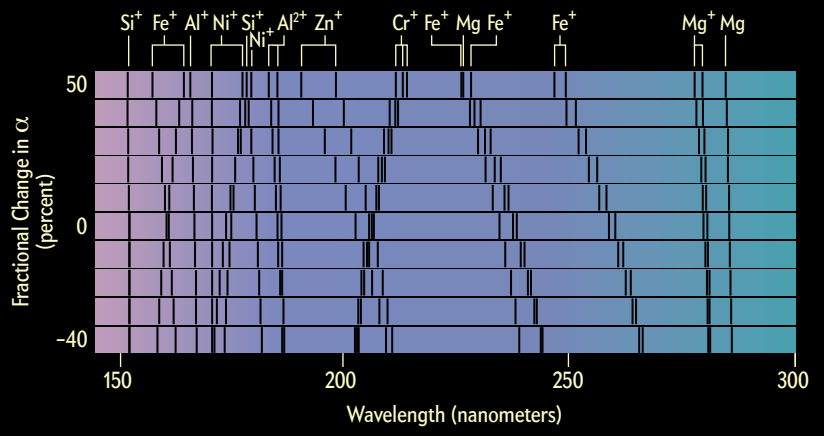
Several of the best-known constants of nature, including the speed of light, can be combined into the fine-structure constant (α)—a number that represents how strongly particles interact through electromagnetic forces. One such interaction is the absorption of photons by atoms. Illuminated by light, an atom absorbs specific colors, each corresponding to photons of a certain wavelength.





Energy levels of electrons within the atom describe the absorption process. The energy of a photon is transferred to an electron, which jumps up the ladder of allowable levels. Each possible jump corresponds to a distinct wavelength. The spacing of levels depends on how strongly the electron is attracted to the atomic nucleus and therefore on α. In the case of magnesium ions (Mg^+), if α were smaller, the levels would be closer together. Photons would need less energy (meaning a longer wavelength) to kick electrons up the ladder.

Simulated spectra show how changing α affects the absorption of near-ultraviolet light by various atomic species. The horizontal black lines represent absorbed wavelengths. Each type of atom or ion has a unique pattern of lines. Changes in the fine-structure constant affect magnesium (Mg), silicon (Si) and aluminum (Al) less than iron (Fe), zinc (Zn), chromium (Cr) and nickel (Ni).



ALISON KENDALL (*illustrations*); SOURCE: JOHN K. WEBB

a cosmic eyeblink. Slow but substantial changes during the long history of the universe would have gone unnoticed.

Fortunately, physicists have found other tests. During the 1970s scientists at the French atomic energy commission noticed something peculiar about the isotopic composition of ore from a uranium mine in Oklo, Gabon: it looked like the waste products of a nuclear reactor. About two billion years ago Oklo must have been the site of a natural reactor.

In 1976 the late Alexander Shlyakhter of the Petersburg Nuclear Physics Institute in Russia and of Harvard University noticed that the ability of a natural reactor to function depends crucially on the precise energy of a particular state of the samarium nucleus that facilitates the capture of neutrons. And that energy depends sensitively on the value of α. So if the fine-structure constant had been slightly different, no chain reaction could have occurred. One did occur, however, which implies that the constant has not changed by more than one part in 10^8 over the past two billion years. (Physicists continue to debate the exact quantitative results because of the inevitable uncertainties about the conditions inside the natural reactor.)

In 1962 P. James E. Peebles and Robert Dicke of Princeton University first applied similar principles to meteorites: the abundance ratios arising from the radioactive decay of different isotopes in these ancient rocks depend on α. The most sensitive constraint involves the beta decay of rhenium into osmium. According to work by Keith Olive of the University of Minnesota, Maxim Pospelov of the University of Victoria in British Columbia and their colleagues, at the time the rocks formed, α was within

© 2018 Scientific American

Looking for Changes in Quasar Light

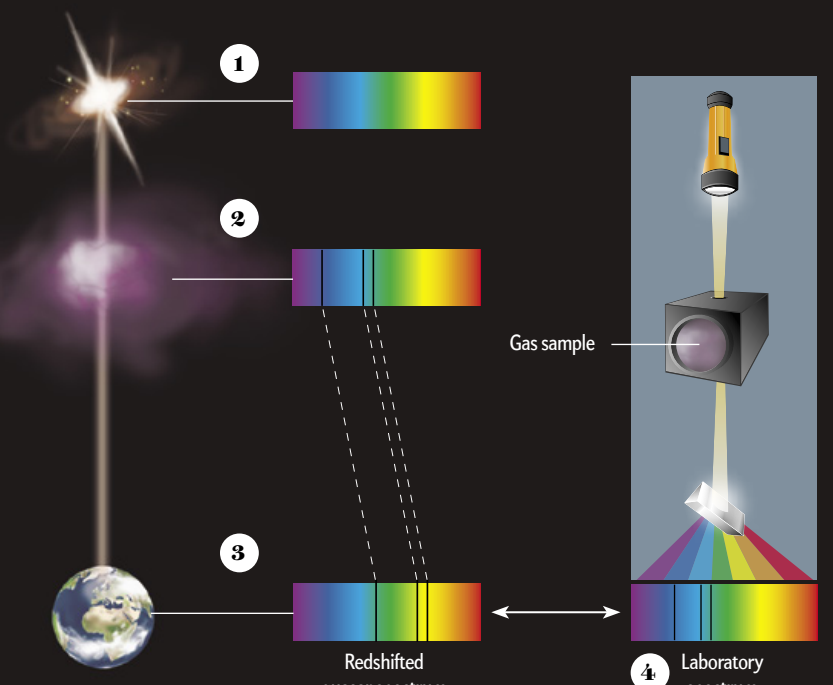
A distant gas cloud, backlit by a quasar, gives astronomers an opportunity to probe the process of light absorption—and therefore the value of the fine-structure constant—earlier in cosmic history.

❶ Light from a quasar begins its journey to Earth billions of years ago with a smooth spectrum.

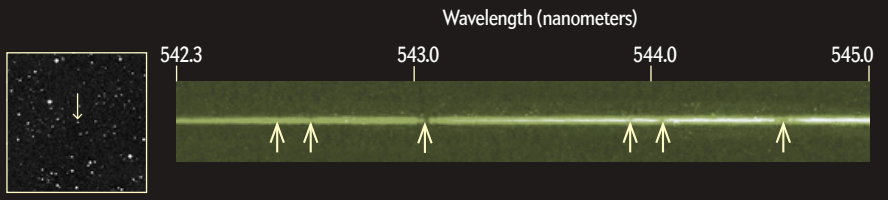
❷ On its way, the light passes through one or more gas clouds. The gas blocks specific wavelengths, creating a series of black lines in the spectrum. For studies of the fine-structure constant, astronomers focus on absorption by metals.

❸ By the time the light arrives on Earth, the wavelengths of the lines have been shifted by cosmic expansion. The amount of shift indicates the distance of the cloud and, hence, its age.

❹ The spacing of the spectral lines can be compared with values measured in the laboratory. A discrepancy suggests that the fine-structure constant used to have a different value.



Quasar spectrum, taken at the European Southern Observatory's Very Large Telescope, shows absorption lines produced by gas clouds between the quasar (*arrow point in square*) and us. The position of the lines (*arrow points on scale*) indicates that the light passed through the clouds about 7.5 billion years ago.



two parts in 10^6 of its current value. This result is less precise than the Oklo data but goes back further in time, to the origin of the solar system 4.6 billion years ago.

To probe possible changes over even longer time spans, researchers must look to the heavens. Light takes billions of years to reach our telescopes from distant astronomical sources. It carries a snapshot of the laws and constants of physics at the time when it started its journey or encountered material en route.

LINE EDITING

ASTRONOMY first entered the constants story soon after the discovery of quasars in 1965. The idea was simple. Quasars had just been discovered and identified as bright sources of light located at huge distances from Earth. Because the path of light from a quasar to us is so long, it inevitably intersects the gaseous outskirts of young galaxies. That gas absorbs the quasar light at particular frequencies, imprinting a bar code of narrow lines onto the quasar spectrum [*see box above*].

Whenever gas absorbs light, electrons within the atoms jump from a low energy state to a higher one. These energy levels are determined by how tightly the atomic nucleus holds the electrons, which depends on the strength of the electromagnetic force between them—and therefore on the fine-structure constant. If the constant was different at the time when the light was absorbed or in the particular region of the universe where it happened, then the energy required to lift the electrons would differ from that required today in lab experiments, and the wavelengths of the transitions seen in the spectra would differ. The way in which the wavelengths change depends critically on the orbital configuration of the electrons. For a given change in α, some wavelengths shrink, whereas others increase. The complex pattern of effects is hard to mimic by data-calibration errors, which makes the test astonishingly powerful.

Before we began our work two decades ago, attempts to perform the measurement had suffered from two limitations. First, lab researchers had not measured the wavelengths of many of the relevant

© 2018 Scientific American

spectral lines with sufficient precision. Ironically, scientists used to know more about the spectra of quasars billions of light-years away than about the spectra of samples here on Earth. We needed some high-precision lab measurements against which to compare the quasar spectra, so we persuaded experimenters to undertake them. Initial measurements were done by Anne Thorne and Juliet Pickering of Imperial College London, followed by groups led by the late Sveneric Johansson of Lund Observatory in Sweden, Ulf Griesmann of the National Institute of Standards and Technology, and Rainer Kling, now at the Karlsruhe Institute of Technology in Germany.

The second problem was that previous observers had used so-called alkali-doublet absorption lines—pairs of absorption lines arising from the same gas, such as carbon or silicon. They compared the spacing between these lines in quasar spectra with lab measurements. This method, however, failed to take advantage of one particular phenomenon: a change in α shifts not just the spacing of atomic energy levels relative to the lowest energy level, or ground state, but also the position of the ground state itself. In fact, this second effect is even stronger than the first. Consequently, the highest precision observers achieved was only about one part in 10^4.

In 1999 one of us (Webb) and Victor V. Flambaum of the University of New South Wales in Sydney came up with a method to take both effects into account. The result was a breakthrough: it meant 10 times higher sensitivity. Moreover, the method allows different species (for instance, magnesium and iron) to be compared, which allows additional cross-checks. Putting this idea into practice took complicated numerical calculations to establish exactly how the observed wavelengths depend on α in all different atom types. Combined with modern telescopes and detectors, the new approach, known as the many-multiplet method, has enabled us to test the constancy of α with unprecedented precision.

CHANGING MINDS

WHEN EMBARKING on this project, we anticipated establishing that the value of the fine-structure constant long ago was the same as it is today; our contribution would simply be higher precision. To our surprise, the first results, in 1999, showed small but statistically significant differences. Further data confirmed this finding. Based on a total of 128 quasar absorption lines, we found an average increase in α of close to six parts in a million over the past six billion to 12 billion years.

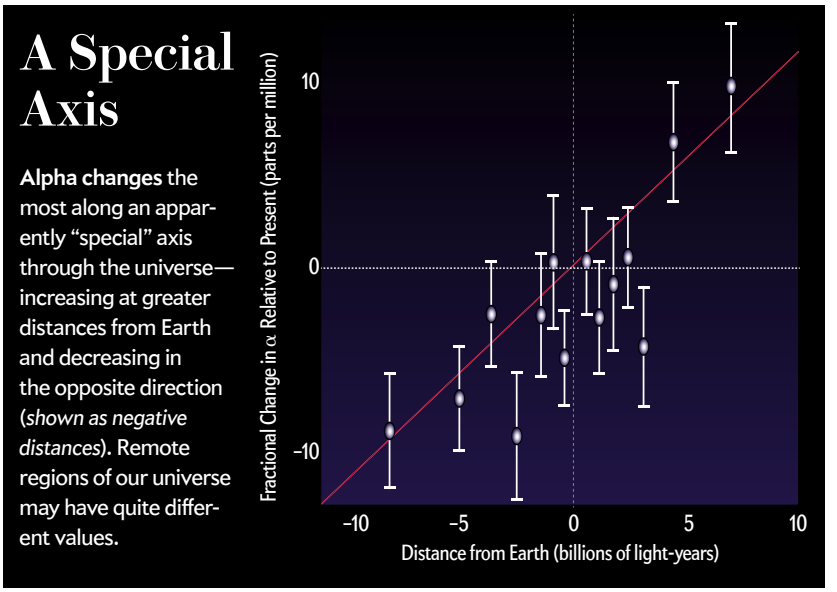
Extraordinary claims require extraordinary evidence, so our immediate thoughts turned to potential problems with the data or the analysis methods. These uncertainties can be classified into two types: systematic and random. Random uncertainties are easier to understand; they are just that—random. They differ for each individual measurement but average out to be close to zero over a large sample. Systematic uncertainties, which do not average out, are harder to deal with. They are endemic in astronomy. Lab experimenters can alter their instrumental setup to minimize them, but astronomers cannot change the universe, and so they are forced to accept that all their methods of gathering data have an irremovable bias. For example, any survey of galaxies will tend to be overrepresented by bright galaxies because they are easier to see. Identifying and neutralizing these biases present a constant challenge.

The first one we looked for was a distortion of the wavelength scale against which the quasar spectral lines were measured. Such a distortion might conceivably be introduced, for example, during the processing of the quasar data from their raw form at the telescope into a calibrated spectrum. Although a simple linear stretching or compression of the wavelength scale could not precisely mimic a change in α, even an imprecise mimicry might be enough to explain our results. To test for problems of this kind, we substituted calibration data for the quasar data and analyzed them, pretending they were quasar data. This experiment ruled out simple distortion errors with high confidence.

For more than two years we put up one potential bias after another, only to rule it out after detailed investigation as too small an effect. So far we have identified just one potentially serious source of bias. It concerns the absorption lines produced by the element magnesium. Each of the three stable isotopes of magnesium absorbs light of a different wavelength, but the three wavelengths are very close to one another, and quasar spectroscopy generally sees the three lines blended as one. Based on lab measurements of the relative abundances of the three isotopes, researchers infer the contribution of each. If these abundances in the young universe differed substantially—as might have happened if the stars that spilled magnesium into their galaxies were, on average, heavier than their counterparts today—those differences could simulate a change in α.

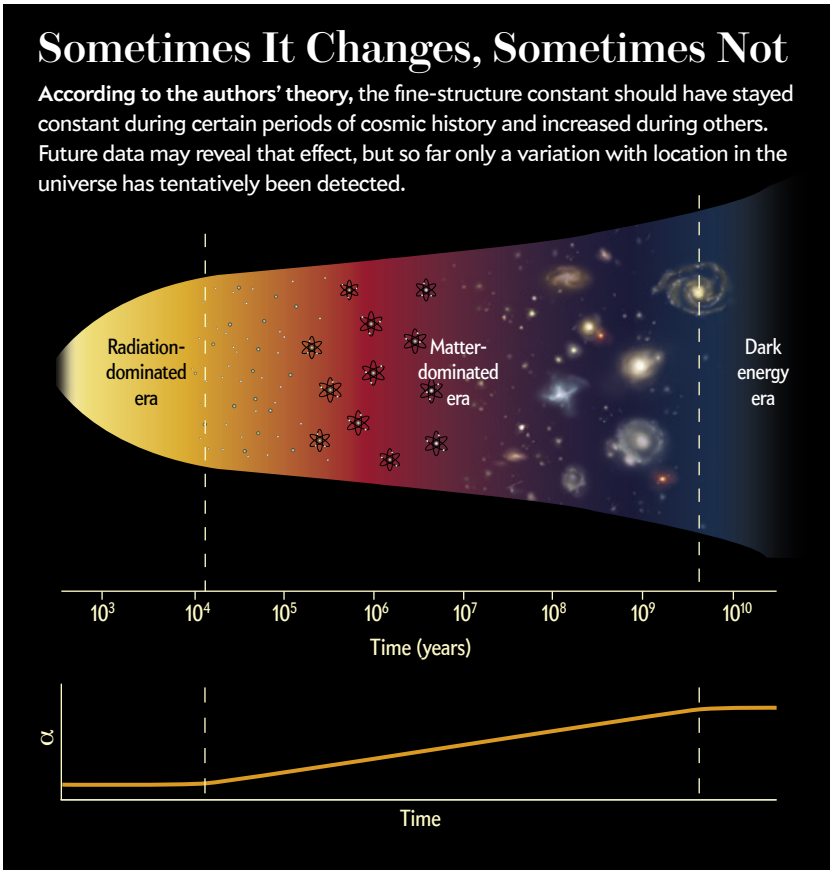
By mid-2010 we completed the analysis of a large amount of new data from the Very Large Telescope (VLT) operated by the European Southern Observatory and obtained 153 new measurements. All the

ALPHA SHIFT


A Special Axis
Alpha changes the most along an apparently "special" axis through the universe—increasing at greater distances from Earth and decreasing in the opposite direction (shown as negative distances). Remote regions of our universe may have quite different values.
Fractional Change in α Relative to Present (parts per million)
10
0
−10
−10
−5
0
5
10
Distance from Earth (billions of light-years)

SCIENTIFIC AMERICAN; SOURCE: JOHN K. WEBB

© 2018 Scientific American


Sometimes It Changes, Sometimes Not
According to the authors' theory, the fine-structure constant should have stayed constant during certain periods of cosmic history and increased during others. Future data may reveal that effect, but so far only a variation with location in the universe has tentatively been detected.
Radiation-dominated era
Matter-dominated era
Dark energy era
10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^10
Time (years)
α
Time

data our group had previously analyzed had come from the Keck telescopes on Mauna Kea in Hawaii. For these new VLT data, everything was different—the telescopes, the spectrograph, the detectors and the software used for the initial stages of the data analysis. These VLT data therefore provided a beautiful cross-check with our results from the Keck telescopes.

We thought it was possible that the new data would show no change in α at all or that they would show the same effect the Keck data did—with α appearing smaller at higher redshifts. What we actually found was truly astonishing and, if correct, will revolutionize some of our most fundamental concepts in physics.

The new VLT data showed not a smaller value of α at high redshift but a *larger* value, larger by just about the same amount as the Keck data are smaller. How can this be? Our immediate thought was that we were seeing evidence for systematic problems in both data sets. Add the Keck and VLT samples together, and to a good approximation, the combined sample shows no change in α with redshift. Problem solved. The constants are really constant after all.

But if that is the explanation, it requires two different systematic effects, one for each telescope, such that both effects are, independently, of the same magnitude but opposite sign. This is not impossible, although so far we have not managed to identify what this unknown pair of systematic effects could be.

We have discovered another curiosity, however. The Keck data cover a largish portion of the sky in the Northern Hemisphere, large enough to ask whether there is any "preferred direction" for the change in α seen with that sample. Put another way: Could it be that α changes not with redshift but with position on the sky? A simple analysis identified one particular direction for which that might be the case. Surprisingly, when the VLT data are analyzed independently, the same direction pops up. The VLT is in Chile and, on average, points to a very different part of the universe than the Keck telescopes do. Another coincidence? Possibly, but that now makes two coincidences.

What happens when we merge the old Keck and the new VLT samples? The result is positively intriguing: the directional dependence becomes highly significant. Deriving such a result by chance appears to be extremely unlikely. If the result is a fluke, we might expect a subset of the data to be generating a rogue result. With this in mind, we devised a simple test to iteratively reduce the sample, discarding one point at a time, to see how much data we needed to eliminate before the apparent spatial dependence of α vanishes. We found that we needed to throw away half the data before the chance probability reduced to a sufficiently unimpressive level! Again, perhaps this is a fluke. Despite extensive attempts, however, we have yet to find a combination of systematic effects in the data that could mimic a spatial dependence. Alpha appears to change spatially—across, perhaps, the entire observable universe. Any change with time is smaller and is currently below our detection sensitivity.

REFORMING THE LAWS

IF OUR FINDINGS prove to be right, the consequences are enormous, though only partially explored. Until quite recently, all attempts to evaluate what happens to the universe if the fine-structure constant changes were unsatisfactory. They amounted to nothing more than assuming that α became a variable in the same formulas that had been derived assuming it was a constant. This is a dubious practice. If α varies, then its effects must conserve energy and momentum, and they must influence the gravitational field in the universe. In 1982 Jacob D. Bekenstein of the Hebrew University of Jerusalem was the first to generalize the laws of electromagnetism to handle inconstant constants rigorously. Bekenstein's theory elevates α from a mere number to a so-called scalar field, a dynamic ingredient of nature. His theory did not include gravity, however. Almost 20 years ago one of us (Barrow), with João Magueijo of Imperial College London, and Håvard B. Sandvik, then also at Imperial, extended it to do so.

This theory makes appealingly simple predictions. Variations in α of a few parts per million should have a completely negligible effect on the expansion of the universe. That is because electromagnetism is much weaker than gravity on cosmic scales. But although changes in the fine-

© 2018 Scientific American

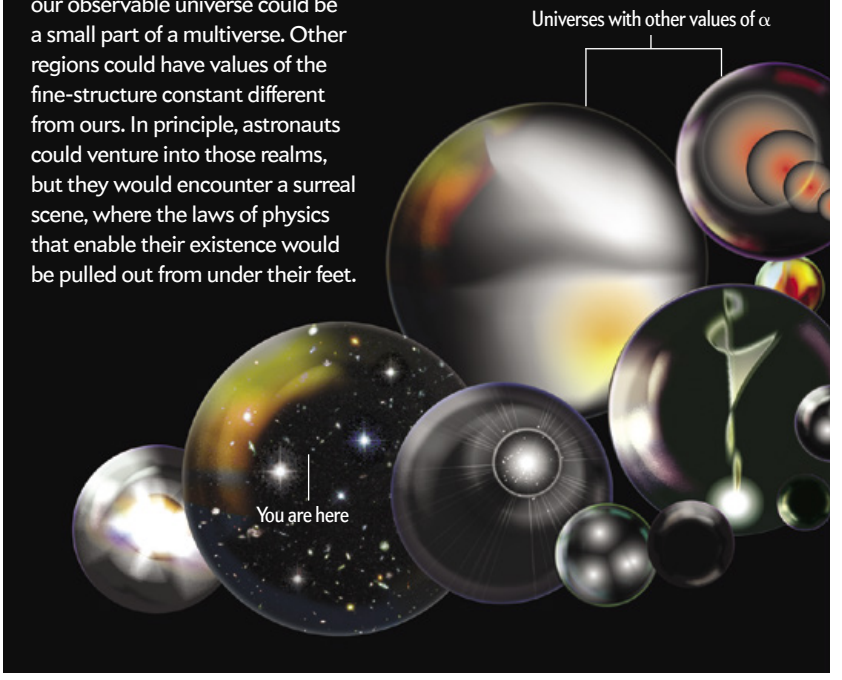
structure constant do not affect the expansion of the universe significantly, the expansion affects α. Changes to α are driven by imbalances between the electric field energy and magnetic field energy. During the first tens of thousands of years of cosmic history, radiation dominated over charged particles and kept the electric and magnetic fields in balance. As the universe expanded, radiation thinned out, and matter became the dominant constituent of the cosmos. The electric and magnetic energies became unequal, and α started to increase very slowly, growing as the logarithm of time. About six billion years ago dark energy took over and accelerated the expansion, making it difficult for all physical influences to propagate through space. So α became nearly constant again.

This predicted pattern was consistent with our earlier data from the Keck telescopes, which seemed to indicate that a redshift dependence of α might be linked to a time variation. But the VLT data throw a large wrench in the works. If the Keck data are right and if the VLT data are also right, even if time variation does take place, it must be small compared with the spatial variation we may now be seeing.

SURREAL WORLDS

Across the Universes

In the grand scheme of things, our observable universe could be a small part of a multiverse. Other regions could have values of the fine-structure constant different from ours. In principle, astronauts could venture into those realms, but they would encounter a surreal scene, where the laws of physics that enable their existence would be pulled out from under their feet.



ALPHA IS JUST THE BEGINNING

ANY THEORY WORTHY of consideration does not merely reproduce observations; it must make novel predictions. The above theory suggests that varying the fine-structure constant makes objects fall differently. Galileo predicted that bodies in a vacuum fall at the same rate no matter what they are made of—an idea known as the weak equivalence principle, which was famously demonstrated when *Apollo 15* astronaut David Scott dropped a feather and a hammer and saw them hit the lunar dirt at the same time. But if α varies, that principle no longer holds exactly. The variations generate a force on all charged particles. The more protons an atom has in its nucleus, the more strongly it will feel this force. If our quasar observations are correct, then the accelerations of different materials differ by about one part in 10^{14}—too small to see in the lab by a factor of about 100 but large enough to show up in planned missions such as STEP (space-based test of the equivalence principle).

So where does this flurry of activity leave science as far as α is concerned? We and many others are eager to either confirm or disprove that α varies at the level claimed. Interestingly, the holdup on new results is not for lack of astronomical data. The Keck and VLT archives already contain plenty of quasar spectra waiting for analysis. The European Southern Observatory has built a new instrument, called ESPRESSO, with higher calibration precision for new measurements of α in the early universe.

If there are plenty of data available with new and better data on the way, what else do we need? The existing measurements took an enormous effort and a great deal of time. The analysis of just one single quasar spectrum is time-consuming and complicated, requires considerable expertise, and involves human—that is, subjective—decision making. To solve this bottleneck, one of us (Webb) worked with then doctoral student Matthew Bainbridge at New South Wales to develop a new artificial-intelligence procedure capable of carrying out fully automated analyses of quasar spectra, producing results that are more objective and reproducible than a human doing the same job. An enormous number of calculations are required for this AI analysis, and supercomputers must be used. That work is now under way, and calculations are being performed on supercomputers in Australia and Britain.

The main focus is on α, over the other constants of nature, simply because it is possible to build up a statistical sample of measurements, mapping the laws of physics throughout the distant cosmos in greater detail. If α is susceptible to change, however, other constants should vary as well, making the inner workings of nature more fickle than scientists ever suspected. SA

MORE TO EXPLORE

Further Evidences for Cosmological Evolution of the Fine Structure Constant. John K. Webb, M. T. Murphy, V. V. Flambaum, V. A. Dzuba, John D. Barrow, C. W. Churchill, J. X. Prochaska and A. M. Wolfe in *Physical Review Letters,* Vol. 87, No. 9, Paper No. 091301; August 27, 2001. http://arxiv.org/abs/astro-ph/0012539

Are the Laws of Nature Changing with Time? John Webb in *Physics World,* Vol. 16, Part 4, pages 33–38; April 2003.

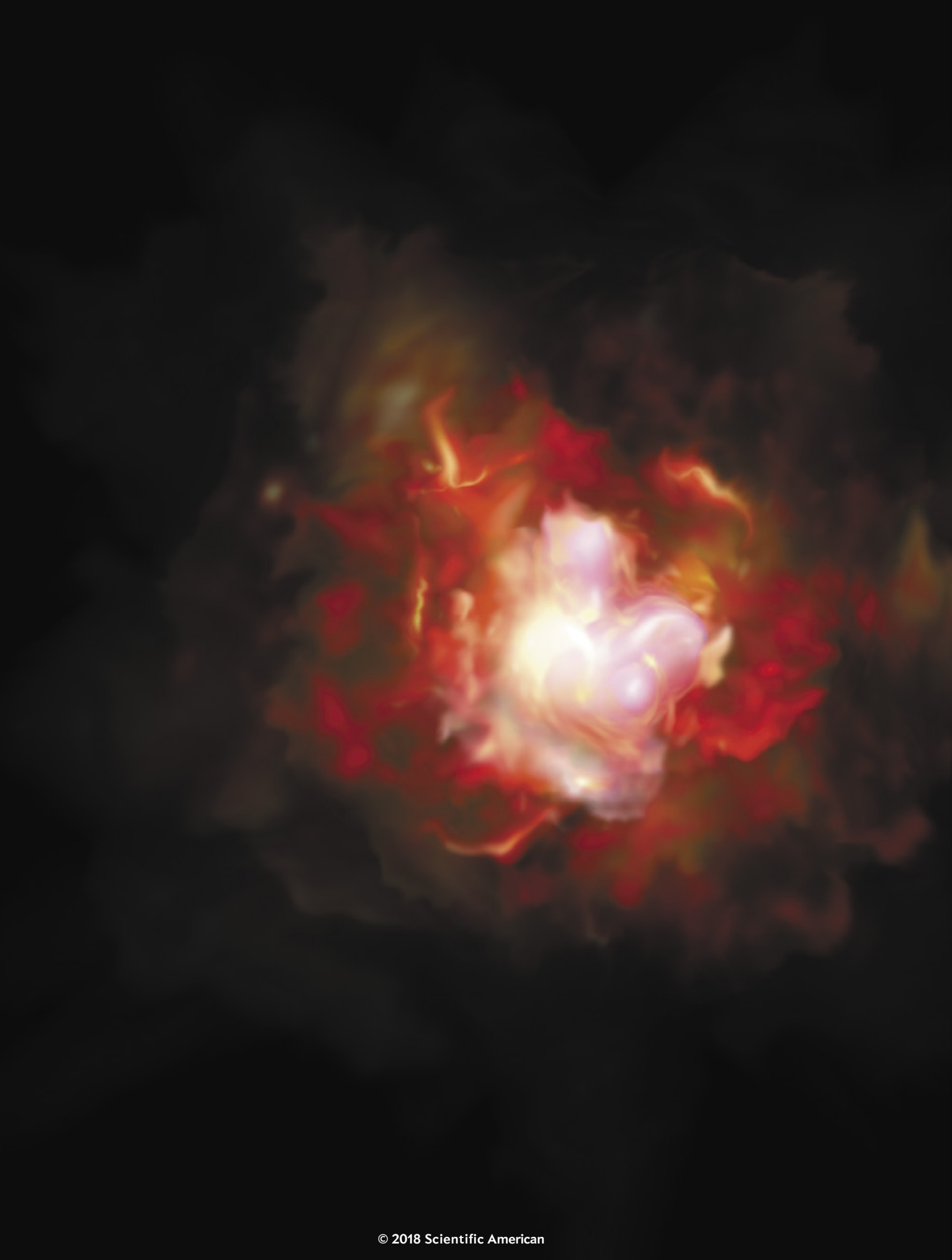
The Constants of Nature: From Alpha to Omega. John D. Barrow. Vintage, 2004.

Indications of a Spatial Variation of the Fine Structure Constant. J. K. Webb, J. A. King, M. T. Murphy, V. V. Flambaum, R. F. Carswell and M. B. Bainbridge in *Physical Review Letters*, Vol. 107, Article No. 191101; October 31, 2011. http://arxiv.org/abs/1008.3907

scientificamerican.com/magazine/sa

RICHARD SWORD

© 2018 Scientific American



© 2018 Scientific American

COSMOLOGY

The Myth of the Beginning of Time

String theory suggests that the big bang was not the origin of the universe but simply the outcome of a preexisting state

By Gabriele Veneziano

Illustration by Alfred T. Kamajian

© 2018 Scientific American

Gabriele Veneziano, a theoretical physicist, formerly at CERN near Geneva, is now an emeritus professor at the College of France in Paris. He was the father of string theory in the late 1960s and one of the first to later apply the theory to black holes and cosmology.

WAS THE BIG BANG REALLY THE BEGINNING OF TIME? OR DID THE universe exist before then? Such a question seemed almost blasphemous only decades ago. Most cosmologists insisted that it simply made no sense—that to contemplate a time before the big bang was like asking for directions to a place north of the North Pole. But developments in theoretical physics, especially the rise of string theory, have changed their perspective. The pre-bang universe has become the latest frontier of cosmology.

The new willingness to consider what might have happened before the bang is the latest swing of an intellectual pendulum that has rocked back and forth for millennia. In one form or another, the issue of the ultimate beginning has engaged philosophers and theologians in nearly every culture. It is entwined with a grand set of concerns, one famously encapsulated in an 1897 painting by Paul Gauguin: *D'ou venons-nous? Que sommes-nous? Ou allons-nous?* "Where do we come from? What are we? Where are we going?" The piece depicts the cycle of birth, life and death—origin, identity and destiny for each individual—and these personal concerns connect directly to cosmic ones. We can trace our lineage back through the generations, back through our animal ancestors, to early forms of life and protolife, to the elements synthesized in the primordial universe, to the amorphous energy deposited in space before that. Does our family tree extend forever backward? Or do its roots terminate? Is the cosmos as impermanent as we are?

The ancient Greeks debated the origin of time fiercely. Aristotle, taking the no-beginning side, invoked the principle that out of nothing, nothing comes. If the universe could never have gone from nothingness to somethingness, it must always have existed. For this and other reasons, time must stretch eternally into the past and future. Christian theologians tended to take the opposite point of view. Augustine contended that God exists outside of space and time, able to bring these

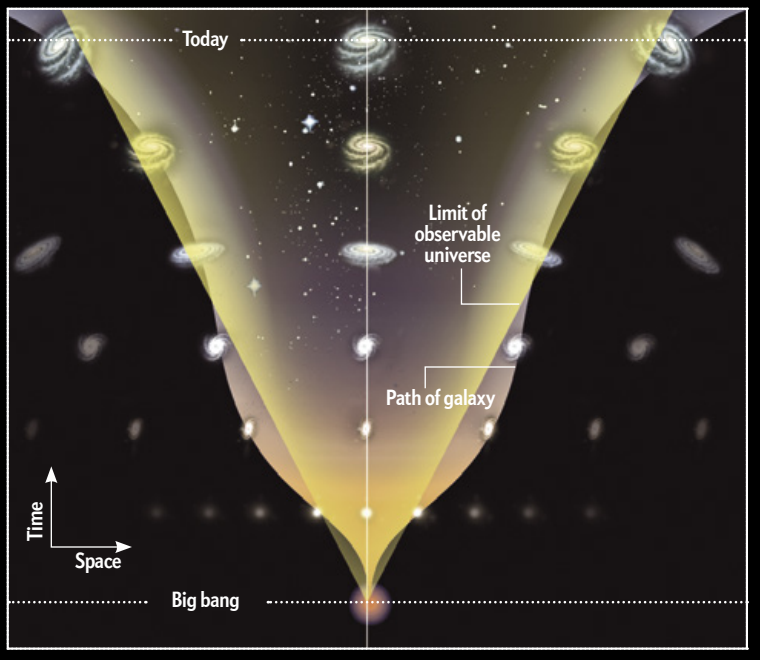
IN BRIEF

Philosophers, theologians and scientists have long debated whether time is eternal or finite—that is, whether the universe has always existed or whether it had a definite genesis. Einstein's general theory of relativity implies finiteness. An expanding universe must have begun at the big bang. **Yet general relativity ceases** to be valid in the vicinity of the bang because quantum mechanics comes into play. The leading candidate for a full quantum theory of gravity—string theory—introduces a minimal quantum of length as a new fundamental constant of nature, making the very concept of a bangian genesis untenable. **The bang still took place,** but it did not involve a moment of infinite density, and the universe may have predated it. The symmetries of string theory suggest that time did not have a beginning and will not have an end. The universe could have begun almost empty and built up to the bang, or it might even have gone through a cycle of death and rebirth.

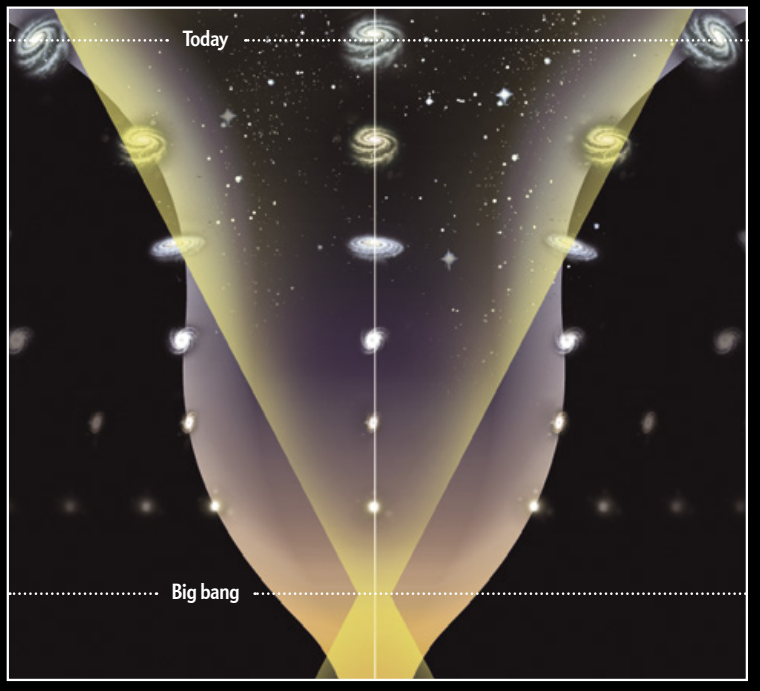
© 2018 Scientific American

Two Views of the Beginning

In our expanding universe, galaxies rush away from one another like a dispersing mob. Any two galaxies recede at a speed proportional to the distance between them: a pair 500 million light-years apart separates twice as fast as one 250 million light-years apart. Thus, all the galaxies we see must have started from the same place at the same time: the big bang. The conclusion holds even though cosmic expansion has gone through periods of acceleration and deceleration; in spacetime diagrams, such as the ones here, galaxies follow sinuous paths that take them in and out of the observable region of space (*yellow wedge*). The situation became uncertain, however, at the precise moment when the galaxies (or their ancestors) began their outward motion.



In standard big bang cosmology, which is based on Einstein's general theory of relativity, the distance between any two galaxies was zero a finite time ago. Before that moment, time loses meaning.



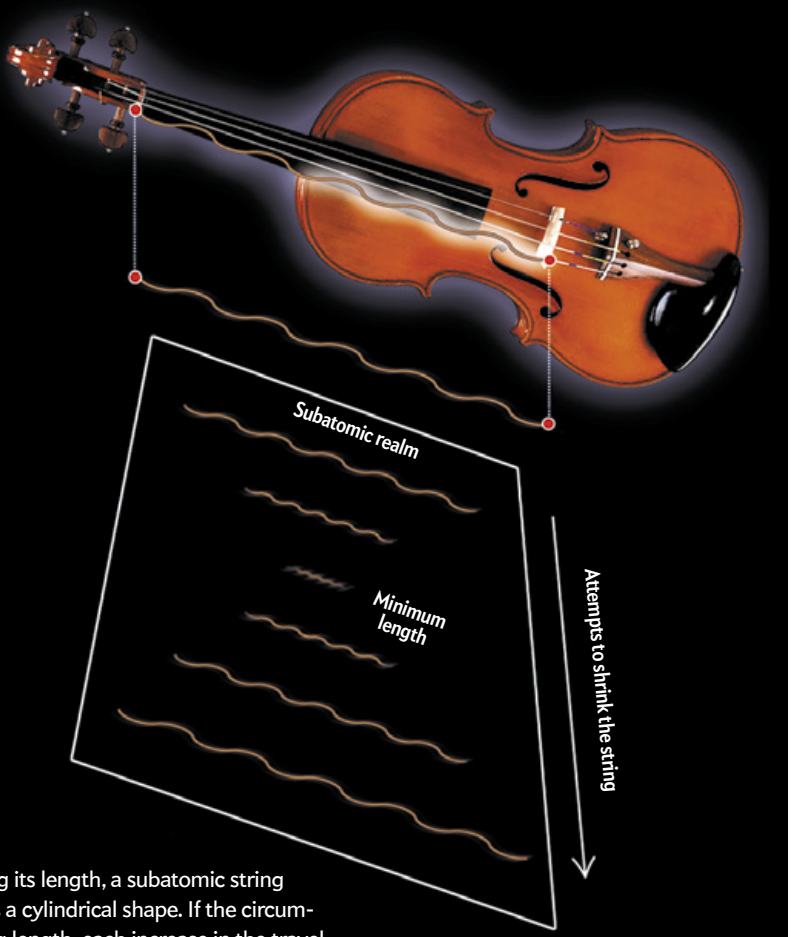
In more sophisticated models, which include quantum effects, any pair of galaxies must have started off a certain minimum distance apart. These models open up the possibility of a pre-bang universe.

Illustrations by Samuel Velasco

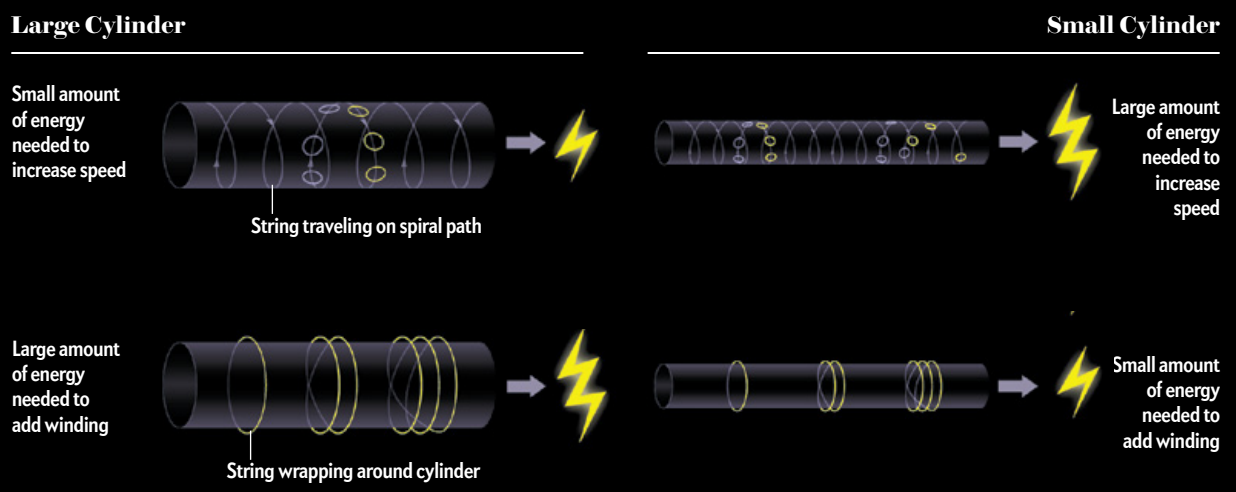
© 2018 Scientific American

String Theory 101

String theory is the leading (though not only) theory that tries to describe what happened at the moment of the big bang. The strings that the theory describes are material objects much like those on a violin. As violinists move their fingers down the neck of the instrument, they shorten the strings and increase the frequency of their vibrations. If they reduced a string to a sub-subatomic length, quantum effects would take over and prevent it from being shortened any further.



In addition to traveling as a unit or vibrating along its length, a subatomic string can wind up like a spring. Suppose that space has a cylindrical shape. If the circumference is larger than the minimum allowed string length, each increase in the travel speed requires a small increment of energy, whereas each extra winding requires a large one. If the circumference is smaller than the minimum length, an extra winding is less costly than an extra bit of velocity. Thus, from a physical point of view, the minimal size of the circumference is not zero but the irreducible quantum of length, symbolized by l_s.


Large Cylinder
Small Cylinder
Small amount of energy needed to increase speed
String traveling on spiral path
Large amount of energy needed to increase speed
Large amount of energy needed to add winding
String wrapping around cylinder
Small amount of energy needed to add winding

© 2018 Scientific American

constructs into existence as surely as he could forge other aspects of our world. When asked, "What was God doing before he created the world?" Augustine answered, "Time itself being part of God's creation, there was simply no before!"

Einstein's general theory of relativity led modern cosmologists to much the same conclusion. The theory holds that space and time are soft, malleable entities. On the largest scales, space is naturally dynamic, expanding or contracting over time, carrying matter like driftwood on the tide. Astronomers confirmed in the 1920s that our universe is currently expanding: distant galaxies move apart from one another. One consequence, as physicists Stephen Hawking and Roger Penrose proved in the 1960s, is that time cannot extend back indefinitely. As you play cosmic history backward in time, the galaxies all come together to a single infinitesimal point, known as a singularity—almost as if they were descending into a black hole. Each galaxy or its precursor is squeezed down to zero size. Quantities such as density, temperature and spacetime curvature become infinite. The singularity is the ultimate cataclysm, beyond which our cosmic ancestry cannot extend.

The unavoidable singularity poses serious problems for cosmologists. In particular, it sits uneasily with the high degree of homogeneity and isotropy that the universe exhibits on large scales. For the cosmos to look broadly the same everywhere, some kind of communication had to pass among distant regions of space, coordinating their properties. Yet the idea of such communication contradicts the old cosmological paradigm.

STRANGE COINCIDENCE

TO BE SPECIFIC, consider what has happened over the 13.8 billion years since the release of the cosmic microwave background radiation. The distance between galaxies has grown by a factor of about 1,000 (because of the expansion), while the radius of the observable universe has grown by the much larger factor of about 100,000 (because light outpaces the expansion). We see parts of the universe today that we could not have seen 13.8 billion years ago. Indeed, this is the first time in cosmic history that light from the most distant galaxies has reached the Milky Way.

Nevertheless, the properties of the Milky Way are basically the same as those of distant galaxies. It is as though you showed up at a party only to find you were wearing exactly the same clothes as a dozen of your closest friends. If just two of you were dressed the same, it might be explained away as coincidence, but a dozen suggests that the partygoers had coordinated their attire in advance. In cosmology, the number is not a dozen but tens of thousands—the number of independent yet statistically identical patches of sky in the microwave background.

One possibility is that all those regions of space were endowed at birth with identical properties—in other words, that the homogeneity is mere coincidence. Physicists, however, have thought about two more natural ways out of the impasse: the early universe was much smaller or much older than in standard cosmology. Either (or both, acting together) would have made intercommunication possible.

IF TIME DID NOT BEGIN AT THE BIG BANG, MATTER COULD HAVE HAD PLENTY OF TIME TO ARRANGE ITSELF SMOOTHLY.

The most popular choice follows the first alternative. It postulates that the universe went through a period of accelerating expansion, known as inflation, early in its history. Before this phase, galaxies or their precursors were so closely packed that they could easily coordinate their properties. During inflation, they fell out of contact because light was unable to keep pace with the frenetic expansion. After inflation ended, the expansion began to decelerate, so galaxies gradually came back into one another's view.

Physicists ascribe the inflationary spurt to the potential energy stored in a new quantum field, the inflaton, about 10^{-35} second after the big bang. Potential energy, as opposed to rest mass or kinetic energy, leads to gravitational repulsion. Rather than slowing down the expansion, as the gravitation of ordinary matter would, the inflaton accelerated it. Proposed in 1981, inflation has explained a wide variety of observations with precision. A number of possible theoretical problems remain, though, beginning with the questions of what exactly the inflaton was and what gave it such a huge initial potential energy.

A less widely known way to solve the puzzle follows the second alternative by getting rid of the singularity. If time did not begin at the bang, if a long era preceded the onset of the present cosmic expansion, matter could have had plenty of time to arrange itself smoothly. Therefore, researchers have reexamined the reasoning that led them to infer a singularity.

One of the assumptions—that relativity theory is always valid—is questionable. Close to the putative singularity, quantum effects must have been important, even dominant. Standard relativity takes no account of such effects, so accepting the inevitability of the singularity amounts to trusting the theory beyond reason. To know what really happened, physicists need to subsume relativity in a quantum theory of gravity. The task has occupied theorists from Albert Einstein onward, but progress was almost zero until the mid-1980s.

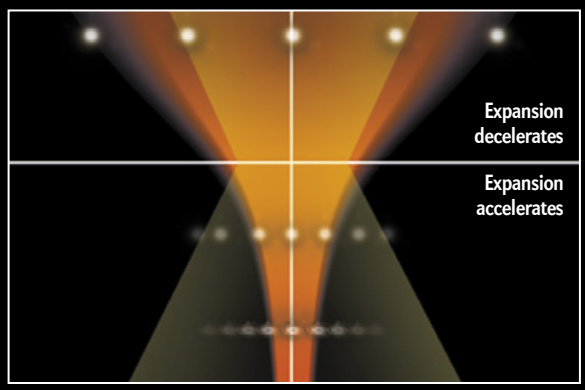
EVOLUTION OF A REVOLUTION

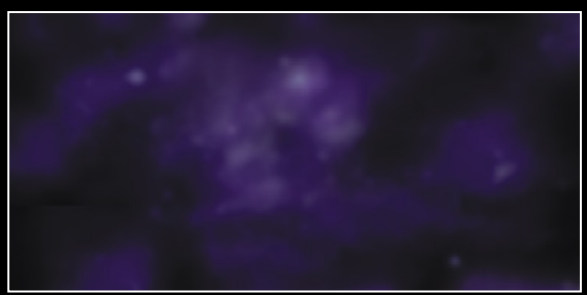
TODAY TWO APPROACHES stand out. One, going by the name of loop quantum gravity, retains Einstein's theory essentially intact but changes the procedure for implementing it in quantum mechanics [see "Atoms of Space and Time," by Lee Smolin, on page 94]. Practitioners of loop quantum gravity have taken great strides and achieved deep insights over the past several years. Still, their approach may not be revolutionary enough to resolve the fundamental problems of quantizing gravity. A similar problem faced particle theorists after Enrico Fermi introduced his effective theory of the weak nuclear force in 1934. All efforts to construct a quantum version of Fermi's theory failed miserably. What was needed was not a

© 2018 Scientific American

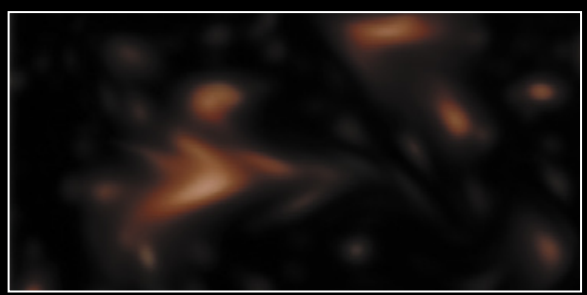
Pre–Big Bang Scenario

A pioneering effort to apply string theory to cosmology was the so-called pre–big bang scenario, according to which the bang is not the ultimate origin of the universe but a transition. Beforehand, expansion accelerated; afterward, it decelerated (at least initially). The path of a galaxy through spacetime is shaped like a wineglass.



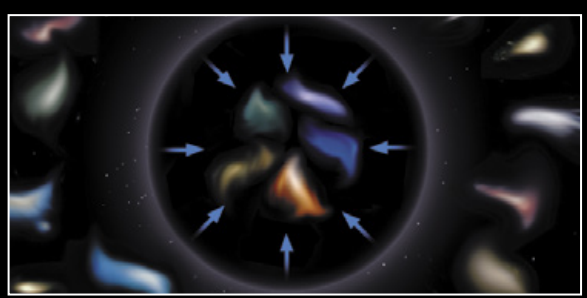


The universe has existed forever. In the distant past it was nearly empty. Forces such as gravitation were inherently weak.

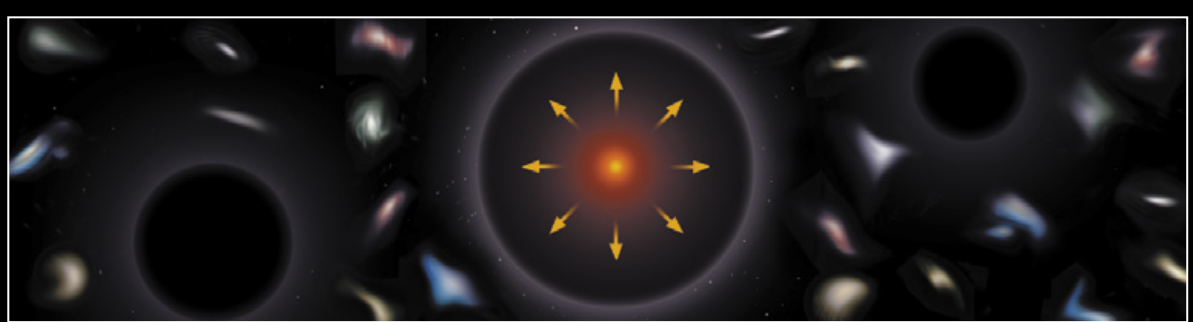


The forces gradually strengthened, so matter began to clump. In some regions, it grew so dense that a black hole formed.

Space inside the hole expanded at an accelerating rate. Matter inside was cut off from matter outside.



Inside the hole, matter fell toward the middle and increased in density until reaching the limit imposed by string theory.



When matter reached the maximum allowed density, quantum effects caused it to rebound in a big bang. Outside, other holes began to form—each, in effect, a distinct universe.

© 2018 Scientific American

new technique but the deep modifications brought by the electroweak theory of Sheldon L. Glashow, Steven Weinberg and Abdus Salam in the late 1960s.

The second approach, which I consider to be more promising, is string theory—a truly revolutionary modification of Einstein's theory. This article will focus on it, although proponents of loop quantum gravity claim to reach many of the same conclusions.

String theory grew out of a model that I wrote down in 1968 to describe the world of nuclear particles (such as protons and neutrons) and their interactions. Despite much initial excitement, the model failed. It was abandoned several years later in favor of quantum chromodynamics, which describes nuclear particles in terms of more elementary constituents: quarks. Quarks are confined inside a proton or a neutron, as if they were tied together by elastic strings. In retrospect, the original string theory had captured those stringy aspects of the nuclear world. Only later was it revived as a candidate for combining general relativity and quantum theory.

The basic idea is that elementary particles are not pointlike but rather infinitely thin, one-dimensional objects—the strings. The large zoo of elementary particles, each with its own characteristic properties, reflects the many possible vibration patterns of a string. How can such a simple-minded theory describe the complicated world of particles and their interactions? The answer can be found in what we may call quantum string magic. Once the rules of quantum mechanics are applied to a vibrating string—just like a miniature violin string, except that the vibrations propagate along it at the speed of light—new properties appear. All have profound implications for particle physics and cosmology.

First, quantum strings have a finite size. Were it not for quantum effects, a violin string could be cut in half, cut in half again, and so on, all the way down, finally becoming a massless, pointlike particle. Heisenberg's uncertainty principle eventually intrudes, however, and prevents the lightest strings from being sliced smaller than about 10^{-34} meter. This irreducible quantum of length, denoted l_s, is a new constant of nature introduced by string theory side by side with the speed of light, c, and Planck's constant, h. It plays a crucial role in almost every aspect of string theory, putting a finite limit on quantities that otherwise could become either zero or infinite.

Second, quantum strings may have angular momentum even if they lack mass. In classical physics, angular momentum is a property of an object that rotates with respect to an axis. The formula for angular momentum multiplies together velocity, mass and distance from the axis; hence, a massless object can have no angular momentum. But quantum fluctuations change the situation. A tiny string can acquire up to two units of h of angular momentum without gaining any mass. This feature is very welcome because it precisely matches the properties of the carriers of all known fundamental forces, such as the photon (for electromagnetism) and the graviton (for gravity). Historically, angular momentum is what clued in physicists to the quantum-gravitational implications of string theory.

Third, quantum strings demand the existence of extra dimensions of space, in addition to the usual three. Whereas a classical violin string will vibrate no matter what the properties of space and time are, a quantum string is more finicky. The equations describing the vibration become inconsistent unless spacetime either is highly curved (in contradiction with observations) or contains six extra spatial dimensions.

Fourth, physical constants—such as Newton's and Coulomb's constants, which appear in the equations of physics and determine the properties of nature—no longer have arbitrary, fixed values. They occur in string theory as fields, rather like the electromagnetic field, that can adjust their values dynamically. These fields may have taken different values in different cosmological epochs or in remote regions of space, and even today the physical "constants" may vary by a small amount. Observing any variation would provide an enormous boost to string theory.

One such field, called the dilaton, is the master key to string theory; it determines the overall strength of all interactions. The dilaton fascinates string theorists because its value can be reinterpreted as the size of an extra dimension of space, giving a grand total of 11 spacetime dimensions.

TYING DOWN THE LOOSE ENDS

FINALLY, quantum strings have introduced physicists to some striking new symmetries of nature known as dualities, which alter our intuition for what happens when objects get extremely small. I have already alluded to a form of duality: typically a short string is lighter than a long one, but if we attempt to squeeze down its size below the fundamental length l_s, the string gets heavier again.

Another form of the symmetry, T-duality, holds that small and large extra dimensions are equivalent. This symmetry arises because strings can move in more complicated ways than pointlike particles can. Consider a closed string (a loop) located on a cylindrically shaped space, whose circular cross section represents one finite extra dimension. Besides vibrating, the string can either turn as a whole around the cylinder or wind around it, one or several times, like a rubber band wrapped around a rolled-up poster [*see box on page 82*].

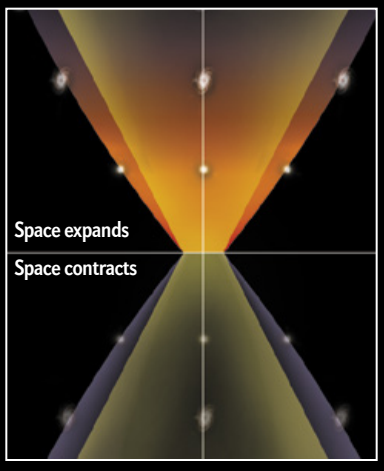
The energetic cost of these two states of the string depends on the size of the cylinder. The energy of winding is directly proportional to the cylinder radius: larger cylinders require the string to stretch more as it wraps around, so the windings contain more energy than they would on a smaller cylinder. The energy associated with moving around the circle, on the other hand, is inversely proportional to the radius: larger cylinders allow for longer wavelengths (smaller frequencies), which represent less energy than shorter wave-

STRINGS ABHOR INFINITY. THEY CANNOT COLLAPSE TO AN INFINITESIMAL POINT, SO THEY AVOID THE PARADOXES THAT COLLAPSE WOULD ENTAIL.

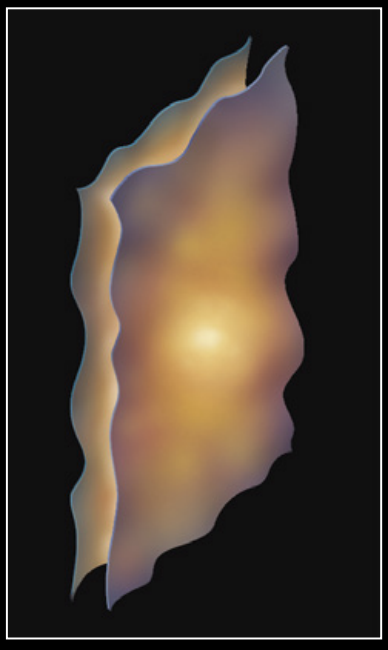
© 2018 Scientific American

Conflagration Scenario

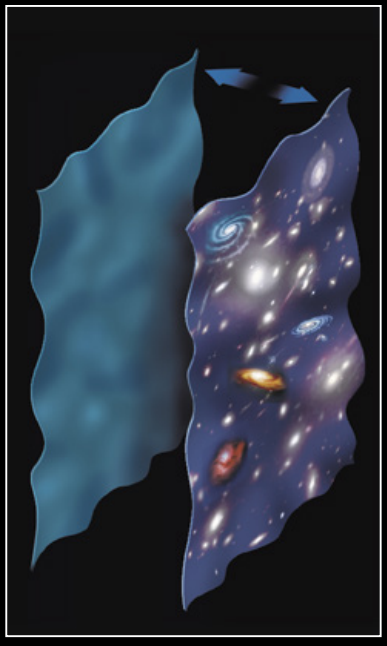
If our universe is a multidimensional membrane, or, simply, a "brane," cruising through a higher-dimensional space, the big bang may have been the collision of our brane with a parallel one. The collisions might recur cyclically. Each galaxy follows an hourglass-shaped path through spacetime.



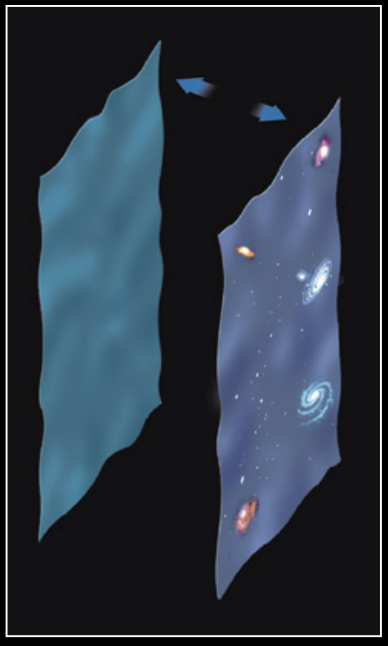
Two nearly empty branes pull each other together. Each is contracting in a direction perpendicular to its motion.



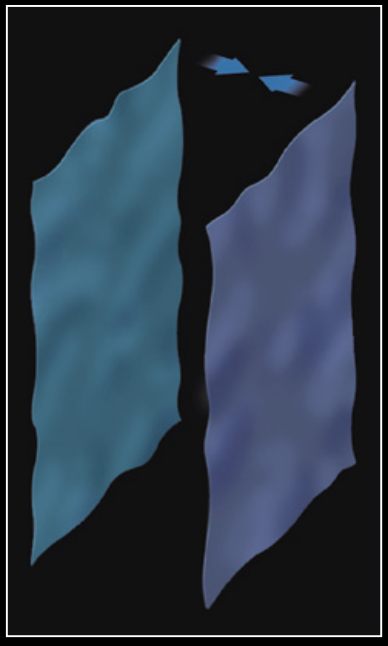
The branes collide, converting their kinetic energy into matter and radiation. This collision is the big bang.



The branes rebound. They start expanding at a decelerating rate. Matter clumps into structures such as galaxy clusters.



In the cyclic model, as the branes move apart, the attractive force between them slows them down. Matter thins out.



The branes stop moving apart and start approaching each other. During the reversal, each brane expands at an accelerated rate.

© 2018 Scientific American

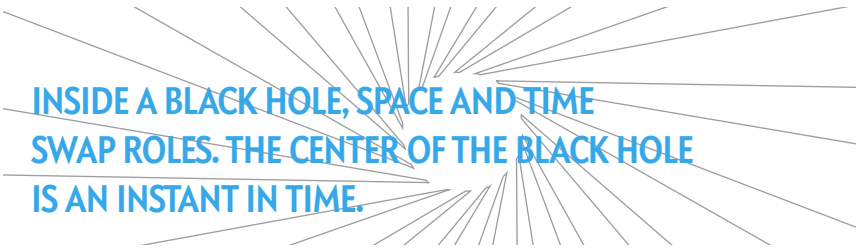
lengths do. If a large cylinder is substituted for a small one, the two states of motion can swap roles. Energies that had been produced by circular motion are instead produced by winding, and vice versa. An outside observer notices only the energy levels, not the origin of those levels. To that observer, the large and small radii are physically equivalent.

Although T-duality is usually described in terms of cylindrical spaces, in which one dimension (the circumference) is finite, a variant of it applies to our ordinary three dimensions, which appear to stretch on indefinitely. One must be careful when talking about the expansion of an infinite space. Its overall size cannot change; it remains infinite. But it can still expand in the sense that bodies embedded within it, such as galaxies, move apart from one another. The crucial variable is not the size of the space as a whole but its scale factor—the factor by which the distance between galaxies changes, manifesting itself as the galactic redshift that astronomers observe. According to T-duality, universes with small scale factors are equivalent to ones with large scale factors. No such symmetry is present in Einstein's equations; it emerges from the unification that string theory embodies, with the dilaton playing a central role.

For years string theorists thought that T-duality applied only to closed strings, as opposed to open strings, which have loose ends and thus cannot wind. In 1995 the late Joseph Polchinski of the University of California, Santa Barbara, realized that T-duality did apply to open strings, provided that the switch between large and small radii was accompanied by a change in the conditions at the end points of the string. Until then, physicists had postulated boundary conditions in which no force acted on the ends of the strings, leaving them free to flap around. Under T-duality, these conditions become so-called Dirichlet boundary conditions, whereby the ends stay put.

Any given string can mix both types of boundary conditions. For instance, electrons may be strings whose ends can move around freely in three of the 10 spatial dimensions but are stuck within the other seven. Those three dimensions form a subspace known as a Dirichlet membrane, or D-brane. In 1996 Petr Horava of the University of California, Berkeley, and Edward Witten of the Institute for Advanced Study in Princeton, N.J., proposed that our universe resides on such a brane. The partial mobility of electrons and other particles explains why we are unable to perceive the full 10-dimensional glory of space.

All the magic properties of quantum strings point in one direction: strings abhor infinity. They cannot collapse to an infinitesimal point, so they avoid the paradoxes that collapse entails. Their nonzero size and novel symmetries set upper bounds to physical quantities that increase without limit in conventional theories, and they set lower bounds to quantities that decrease. String theorists expect that when one plays the history of the universe backward in time, the curvature of spacetime starts to increase. But instead of going all the way to infinity (at the traditional big bang singularity), it eventually hits a maximum and shrinks once more. Before string theory, physicists were hard-pressed to imagine any mechanism that could so cleanly eliminate the singularity.



TAMING THE INFINITE

CONDITIONS NEAR the zero time of the big bang were so extreme that no one yet knows how to solve the equations. Nevertheless, string theorists have hazarded guesses about the pre-bang universe. Two popular models are floating around.

The first, known as the pre–big bang scenario, which my colleagues and I began to develop in 1991, combines T-duality with the better-known symmetry of time reversal, whereby the equations of physics work equally well when applied backward and forward in time. The combination gives rise to new possible cosmologies in which the universe, say, five seconds before the big bang expanded at the same pace as it did five seconds after the bang. Yet the rate of change of the expansion was opposite at the two instants: if it was decelerating after the bang, it was accelerating before. In short, the big bang may not have been the origin of the universe but simply a violent transition from acceleration to deceleration.

The beauty of this picture is that it automatically incorporates the great insight of standard inflationary theory—namely, that the universe had to undergo a period of acceleration to become so homogeneous and isotropic. In the standard theory, acceleration occurs after the big bang because of an ad hoc inflaton field. In the pre–big bang scenario, it happens before the bang as a natural consequence of the novel symmetries of string theory.

According to the scenario, the pre-bang universe was almost a perfect mirror image of the post-bang one [*see box on page 84*]. If the universe is eternal into the future, its contents thinning to a meager gruel, it is also eternal into the past. Infinitely long ago it was nearly empty, filled only with a tenuous, widely dispersed, chaotic gas of radiation and matter. The forces of nature, controlled by the dilaton field, were so feeble that particles in this gas barely interacted.

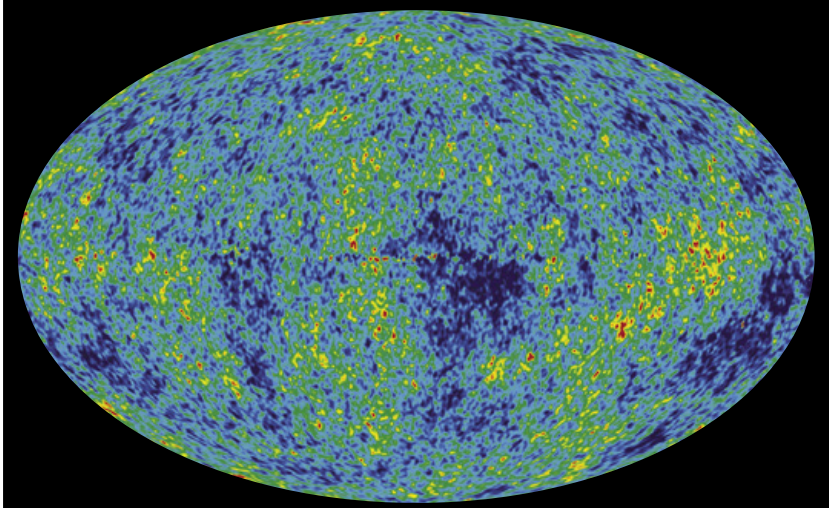
As time went on, the forces gained in strength and pulled matter together. Randomly, some regions accumulated matter at the expense of their surroundings. Eventually the density in these regions became so high that black holes started to form. Matter inside those regions was then cut off from the outside, breaking up the universe into disconnected pieces.

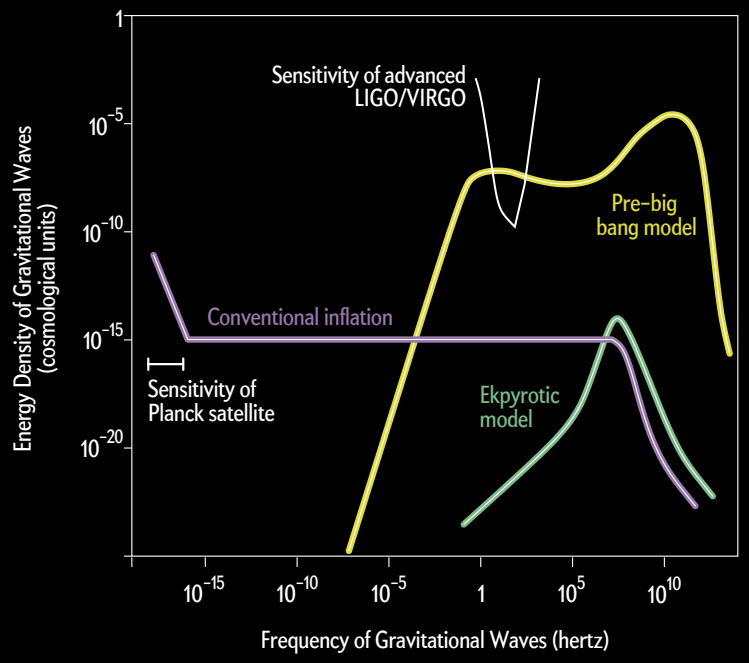
Inside a black hole, space and time swap roles. The center of the black hole is not a point in space but an instant in time. As the infalling matter approached the center, it reached higher and higher densities. But when the density, temperature and curvature reached the maximum values allowed by string theory, these quantities bounced and started decreasing. The moment of that reversal, called a big bang, was later renamed a "big bounce." The interior of one of those black holes became our universe.

© 2018 Scientific American

Observations

Observing the pre-bang universe may sound like a hopeless task, but one form of radiation could survive from that epoch: gravitational radiation. These periodic variations in the gravitational field might be detected indirectly by their effect on the polarization of the cosmic microwave background (*simulated view*) or directly at ground-based observatories. The pre–big bang and ekpyrotic scenarios predict more high-frequency gravitational waves and fewer low-frequency ones than do conventional models of inflation (*shown in graph*). Existing measurements of various astronomical phenomena cannot distinguish among these models, but observations by the Planck satellite, as well as the LIGO and VIRGO observatories, should be able to.





Not surprisingly, such an unconventional scenario has provoked controversy. Andrei Linde of Stanford University has argued that for this scenario to match observations, the black hole that gave rise to our universe would have to have formed with an unusually large size—much larger than the length scale of string theory. An answer to this objection is that the equations predict black holes of all possible sizes. Our universe just happened to form inside a sufficiently large one.

A more serious objection, raised by Thibault Damour of the Institute of Advanced Scientific Studies in Bures-sur-Yvette, France, and Marc Henneaux of the Free University of Brussels, is that matter and spacetime would have behaved chaotically near the moment of the bang, in possible contradiction with the observed regularity of the early universe. I have proposed that a chaotic state would produce a dense gas of miniature "string holes"—strings that were so small and massive that they were on the verge of becoming black holes. The behavior of these holes could solve the problem identified by Damour and Henneaux. A similar proposal has been put forward by Thomas Banks, now at the University of California, Santa Cruz, and Willy Fischler of the University of Texas at Austin. Other critiques also exist, and whether they have uncovered a fatal flaw in the scenario remains to be determined.

The other leading model for the universe before the big bang is the so-called ekpyrotic ("conflagration") scenario. Developed by a team of cosmologists and string theorists—Justin Khoury, now at the University of Pennsylvania, Paul J. Steinhardt of Princeton University, Burt A. Ovrut of the University of Pennsylvania, Nathan Seiberg of the Institute for Advanced Study and Neil Turok, now at the Perimeter Institute for Theoretical Physics in Ontario—the ekpyrotic scenario relies on the previously mentioned Horava-Witten idea that our universe sits at one end of a higher-dimensional space and a "hidden brane" sits at the opposite end. The two branes exert an attractive force on each other and occasionally collide, making the extra dimension shrink to zero before growing again. The big bang would correspond to the time of collision [*see box on page 86*].

In a variant of this scenario, the collisions occur cyclically. Two branes might hit, bounce off each other, move apart, pull

COURTESY OF NASA/WMAP SCIENCE TEAM (*map*)

© 2018 Scientific American

each other together, hit again, and so on. In between collisions, the branes behave like Silly Putty, expanding as they recede and contracting somewhat as they come back together. During the turnaround, the expansion rate accelerates; indeed, the present accelerating expansion of the universe may augur another collision.

The pre–big bang and ekpyrotic scenarios share some common features. Both begin with a large, cold, nearly empty universe, and both share the difficult (and unresolved) problem of making the transition between the pre- and the post-bang phase. Mathematically, the main difference between the scenarios is the behavior of the dilaton field. In the pre–big bang, the dilaton begins with a low value—so that the forces of nature are weak—and steadily gains strength. The opposite is true for the ekpyrotic scenario, in which the collision occurs when forces are at their weakest.

The developers of the ekpyrotic theory initially hoped that the weakness of the forces would allow the bounce to be analyzed more easily, but they were still confronted with a difficult high-curvature situation, so the jury is out on whether the scenario truly avoids a singularity. Also, the ekpyrotic scenario must entail very special conditions to solve the usual cosmological puzzles. For instance, the about-to-collide branes must have been almost exactly parallel to one another, or else the collision could not have given rise to a sufficiently homogeneous bang. The cyclic version may be able to take care of this problem because successive collisions would allow the branes to straighten themselves.

Leaving aside the difficult task of fully justifying these two scenarios mathematically, physicists must ask whether they have any observable physical consequences. At first sight, both scenarios might seem like an exercise not in physics but in metaphysics—interesting ideas that observers could never prove right or wrong. That attitude is too pessimistic. Like the details of the inflationary phase, those of a possible pre-bangian epoch could have observable consequences, especially for the small variations observed in the cosmic microwave background temperature.

First, observations show that the temperature fluctuations were shaped by acoustic waves for several hundred thousand years. The regularity of the fluctuations indicates that the waves were synchronized. Cosmologists have discarded many cosmological models over the years because they failed to account for this synchrony. The inflationary, pre–big bang and ekpyrotic scenarios all pass this first test. In these three models, the waves were triggered by quantum processes amplified during the period of accelerating cosmic expansion. The phases of the waves were aligned.

Second, each model predicts a different distribution of the temperature fluctuations with respect to angular size. Observers have found that fluctuations of all sizes have approximately the same amplitude. (Discernible deviations occur only on very small scales, for which the primordial fluctuations have been altered by subsequent processes.) Inflationary models neatly reproduce this distribution. During inflation, the curvature of space changed relatively slowly, so fluctuations of different sizes were generated under much the same conditions. In both the stringy models, the curvature evolved quickly, increasing the amplitude of small-scale fluctuations, but other processes boosted the large-scale ones, leaving all fluctuations with the same strength. For the ekpyrotic scenario, those other processes involved the extra dimension of space, the one that separated the colliding branes. For the pre–big bang scenario, they involved a quantum field, the axion, related to the dilaton. In short, all three models match the data.

Third, variations in temperature can arise from two distinct processes in the early universe: fluctuations in the density of matter and rippling caused by gravitational waves. Inflation involves both processes, whereas the pre–big bang and ekpyrotic scenarios involve density variations for the most part. Gravitational waves of certain sizes would leave a distinctive signature in the polarization of the cosmic microwave background. Satellite and ground-based observatories may be able to see that signature, if it exists—providing a nearly definitive test.

A fourth test pertains to the statistics of the fluctuations. In inflation the fluctuations follow a bell-shaped curve, which is known to physicists as a Gaussian. The same may be true in the ekpyrotic case, whereas the pre–big bang scenario allows for sizable deviation from Gaussianity.

Analysis of the microwave background is not the only way to verify these theories. The pre–big bang scenario should also produce a random background of gravitational waves in a range of frequencies that, though irrelevant for the microwave background, should be detectable by future gravitational-wave observatories. Moreover, because the pre–big bang and ekpyrotic scenarios involve changes in the dilaton field, which is coupled to the electromagnetic field, they would both lead to large-scale magnetic field fluctuations. Vestiges of these fluctuations might show up in galactic and intergalactic magnetic fields.

VESTIGES OF THE PRE-BANGIAN EPOCH MIGHT SHOW UP IN GALACTIC AND INTERGALACTIC MAGNETIC FIELDS.

So when did time begin? Science does not have a conclusive answer yet, but at least two potentially testable theories plausibly hold that the universe—and therefore time—existed well before the big bang. If either scenario is right, the cosmos has always been in existence and, even if it recollapses one day, will never end. SA

MORE TO EXPLORE

Pre-Big-Bang in String Cosmology. Maurizio Gasperini and Gabriele Veneziano in *Astroparticle Physics*, Vol. 1, No. 3, pages 317–339; July 1993. http://arxiv.org/abs/hep-th/9211021

Superstring Cosmology. James E. Lidsey, David Wands and Edmund J. Copeland in *Physics Reports*, Vol. 337, Nos. 4–5, pages 343–492; October 2000. http://arxiv.org/abs/hep-th/9909061

From Big Crunch to Big Bang. Justin Khoury, Burt A. Ovrut, Nathan Seiberg, Paul J. Steinhardt and Neil Turok in *Physical Review D*, Vol. 65, No. 8, Paper No. 086007; April 15, 2002. http://arxiv.org/abs/hep-th/0108187

A Cyclic Model of the Universe. Paul J. Steinhardt and Neil Turok in *Science*, Vol. 296, pages 1436–1439; May 24, 2002. http://arxiv.org/abs/hep-th/0111030

© 2018 Scientific American

Q&A

INTERVIEW

What Keeps Time Moving Forward?

A timely conversation with physicist Sean M. Carroll about how our one-way trip from the past to the future is entangled with entropy and the origin of the universe

By John Matson

IN BRIEF

WHO
SEAN M. CARROLL

VOCATION/AVOCATION
Research professor in physics, studying the structure and evolution of the universe

WHERE
California Institute of Technology

RESEARCH FOCUS
Cosmology, field theory, particle physics and gravitation

BIG PICTURE
Carroll wants to know why time moves in one direction and what happened at or before the big bang.

PHYSICISTS OFTEN DESCRIBE THE FABRIC OF THE UNIVERSE WE INHABIT AS four-dimensional spacetime, comprising three dimensions of space and one of time. Yet whereas we spend our days passing freely through space in any direction we wish (gravity and solid obstacles permitting), time pushes us along, willingly or not, in a single, predetermined direction: toward the future.

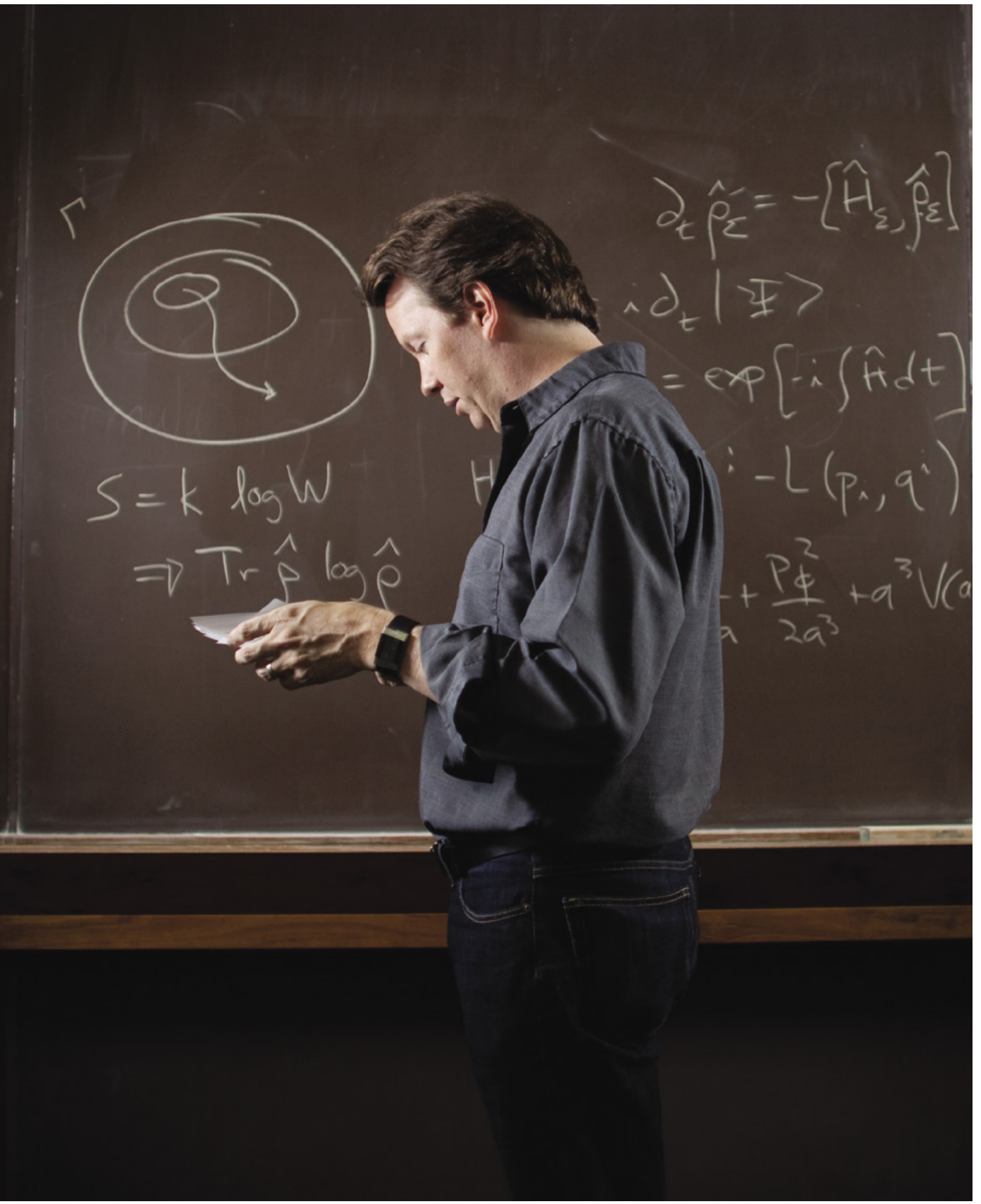
This is the arrow of time—life carries us from the past, through the present, and into the future. *Back to the Future* plotlines notwithstanding, no one knows how to reverse the arrow—how to move backward in time—and the logical paradoxes that would result from such a trip into the past render it a thorny proposition at best. (Thanks to a prediction of special relativity called time dilation, travel into the distant future is relatively easy: just move really, really fast.)

In his 2010 book, *From Eternity to Here,* theoretical physicist Sean M. Carroll of the California Institute of Technology sets out to explain why time marches along unfailingly in one direction. Carroll argues for the necessity of marrying three seemingly disparate concepts: time, entropy and cosmology.

Entropy, which, in rough terms, is the measure of a system's disorder, creeps up over time, as dictated by the second law of thermodynamics. To illustrate entro-

ROBERT GALLAGHER

© 2018 Scientific American



© 2018 Scientific American

py's inexorable growth, Carroll takes us to the breakfast table—you can't unscramble an egg, he points out, and you can't unstir the milk out of your coffee. These systems invariably proceed to disordered, or high-entropy, arrangements.

THE FACT THAT THE PAST IS SET IN STONE WHILE THE FUTURE CAN STILL BE ALTERED IS ALL BECAUSE OF ENTROPY.

Each of these examples shows how the continual growth of entropy fills the world with irreversible processes that divide the past from the future: the making of an omelet and the mixing of milk into a cup of coffee are events that work in only one temporal direction.

But why should entropy always increase? This is where Carroll turns to cosmology, which must explain what happened at the hot, dense big bang and, even before that, when the universe began.

We spoke to the physicist about his book and the challenges of presenting cutting-edge physics to a wide audience. Carroll sees the present as a particularly exciting time to be a theoretical physicist. There is a flood of new data in cosmology, and new experiments are under way at particle accelerators and elsewhere.

What's so interesting about time? To a naive observer, it's something that just passes by and that we can't do anything with; it's unchanging.

There are two things that inspired me to write this book. One is that time is something we all are familiar with. We all use it—we have no problem reading a watch. But then, when you act like a good scientist or philosopher and try to make sense of it, this puzzle arises: the fundamental laws of physics treat the past and the future as being exactly the same, whereas the world does not. There's a big difference—the past happened, and the future is still up for grabs. So it would be nice to know how to reconcile that. That's the arrow of time problem as it's been thought about for at least a couple of hundred years.

I think this is an important and interesting problem, and it's just as good to write about as anything else. But there is something that I think makes this problem a little bit special, which is that the answer to why the past is different from the future, whatever it is going to end up being, is not just about what happens here as you and I are talking, as time goes by in our daily lives. It is intimately connected with the whole universe—with what happened at the big bang, with the special condition in our universe when it started.

A full understanding of what happens in our everyday lives needs to take into account what happened at the big bang. Not only is that intrinsically interesting and just kind of cool to think about, it's also a mystery that is not given much attention by working scientists; it's a little bit underappreciated. We are so far from knowing what the final answer to this is that we sort of don't think about it that much. So I wanted to draw attention to this connection between the arrow of time and cosmology, both to everyday readers and to my scientific friends. I think this is something that we really should keep in mind as one of the fundamental puzzles facing us in modern science.

As an everyday reader, I appreciated the introductory quotes to the chapters from* Annie Hall, *Vladimir Nabokov,* Dumb and Dumber. *How much of a challenge was it to try to keep this book accessible and enjoyable?

I tried my best, and I think I succeeded in some places more than others. A lot of the material was not exactly what I do research on, so I had to learn a lot and sort of think about things I had been vaguely aware of for awhile. I actually think that I was better at making those sections lively and interesting and accessible than the sections that I understood the best. Because I knew I had to sit down and think very, very hard about it; I couldn't just give my conventional spiel.

The good news is that, except for a few things about quantum mechanics and the multiverse, most of the basic ideas are pretty graspable. They're not dramatically abstract; we're not working in higher dimensions or anything like that. You can see the basic ideas we're talking about working themselves out in everyday life.

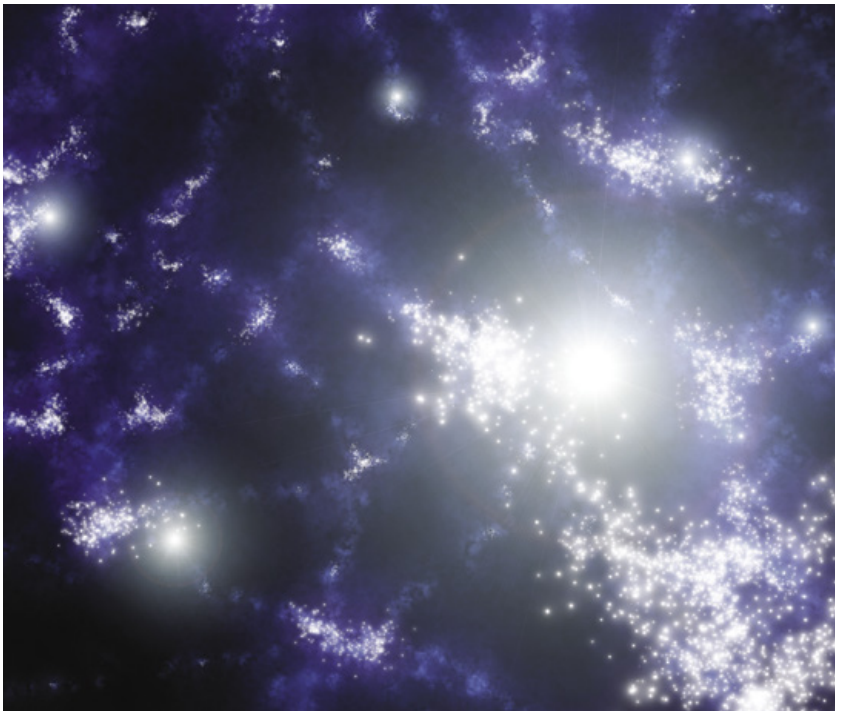
I'm a big believer that science is part of a larger cultural thing. Science is not all by itself. So I definitely wanted to give the feeling that as we're thinking about the universe, and space and time, and experience, and memory, and free will, and all these things that I talk about in the book, this is science, and our everyday lives, and the culture in which we live, so why not sort of have fun and bring them together?

Going back to the science for a moment, how does the concept of entropy intertwine with the arrow of time?

I think people have probably heard the word "entropy." It goes up; that's the second law of thermodynamics. There's this famous—at least famous among scientists—episode with [English novelist and physicist] C. P. Snow, where he was trying to convince people that they should not only be literate about literature but literate about science as well. And the one example Snow chose as something everyone should know was the second law of thermodynamics, the law that says that entropy increases.

And that's true, and I believe that's a great example, but what I think is actually underappreciated is that just about everything about the arrow of time—what we would think of as "how time works," the fact that the past is set in stone while the future can still be altered—is all because of entropy. The fact that you can remember yesterday but not tomorrow is because of entropy. The

© 2018 Scientific American



Starburst: This artist's concept shows the universe just after its first stars were born.

NASA/JPL-CALTECH AND R. HURT SSC

fact that you're always born young and then you grow older, and not the other way around like in *Benjamin Button*—it's all because of entropy. So I think that entropy is underappreciated as something that has a crucial role in how we go through life.

You have a sort of candid moment in the book where you talk about a veteran, unnamed physicist who had some complaints with your theories about time and the second law.
His objection was with the idea that cosmology has something to do with it.

The following statement is very true: To understand the second law of thermodynamics, or how the arrow of time works in our everyday lives, we don't need to ever talk about cosmology. If you pick up a textbook on statistical mechanics, there will be no talk about cosmology at all. So it would be incorrect to say that we need to understand the big bang to use the second law of thermodynamics, to know how it works. The problem is this: to understand why it exists at all requires a knowledge of cosmology and what happened at the big bang.

Once you assume that the universe had a low entropy for whatever reason, everything else follows, and that's all we ever talk about in textbooks. But we're being a little bit more ambitious than that. We want to understand why it was that way—why was it that the entropy was lower yesterday than it is today?

To understand why the entropy was lower yesterday really requires cosmology. If you sit down and think about it carefully, there is absolutely no question that that is true, although a lot of people don't quite accept it yet.

If you take this approach and look at time from a cosmological standpoint, what is this low-entropy condition in the past? What does that look like?
We're not learning anything about the early universe by making this observation. We already know what the early universe was like. It was smooth; it was expanding very rapidly; it was a dense, hot state; and there was a lot of stuff in it. Now, that happens to be a very low entropy configuration that the universe could be in, and that is the puzzle. So it's not that we're learning what the early universe was like, because we already knew that—it's that in trying to explain it, in trying to come up with a theory, whether it's inflation or the cyclic universe or a big bounce, you haven't succeeded in explaining the early universe unless you've explained why it has low entropy. I just think a tremendous number of contemporary cosmological theories fail at that requirement; they sort of sidestep their way around that question rather than addressing it head-on.

Do these various theories make predictions we could test based on our understanding of time and entropy?
Not yet. We'd like them to. All I can say is I hope they will. I talk about that in the epilogue of the book.

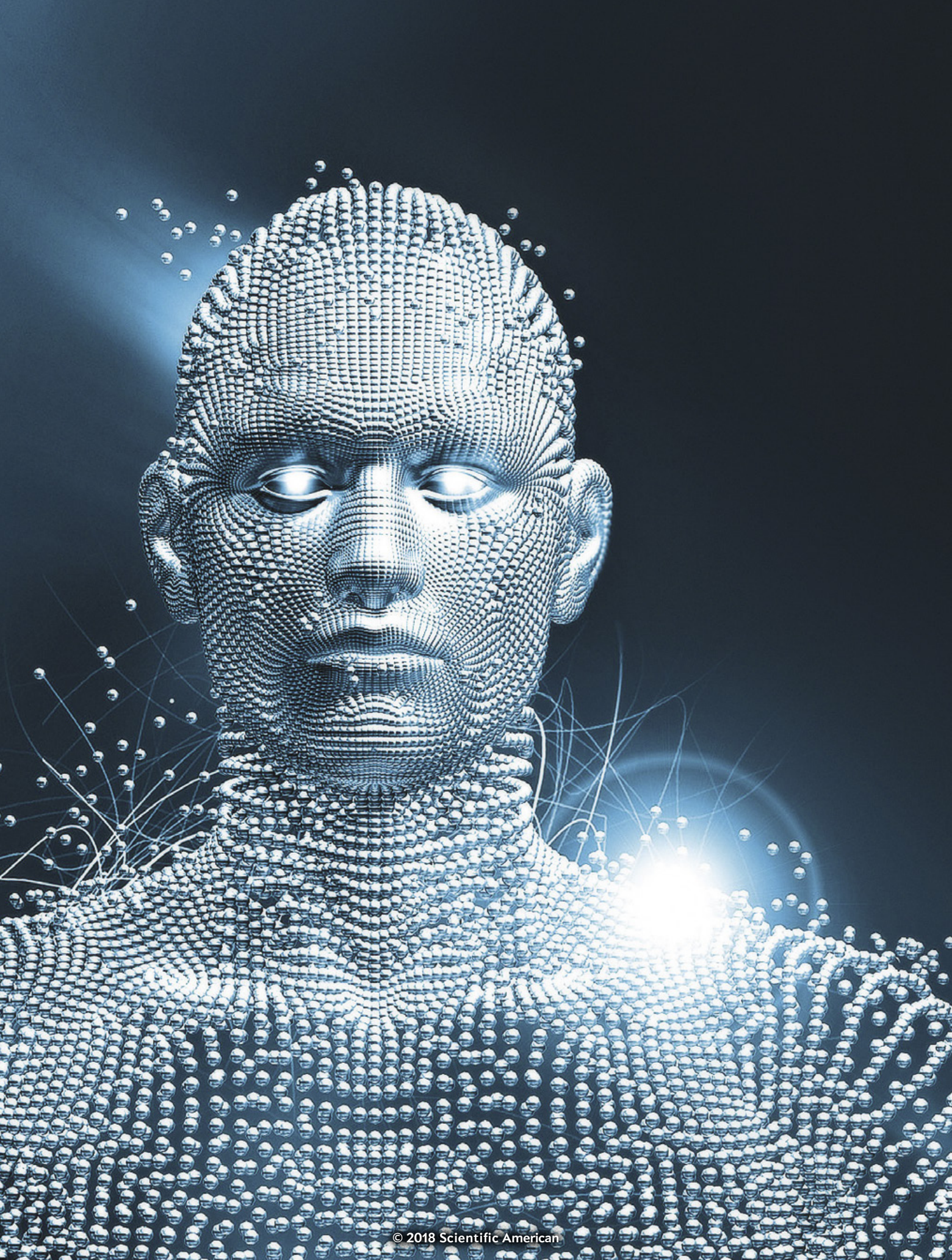
On the one hand, if these ideas don't connect with observed things, then there's no use talking about them. But that's not the same as saying that because we can't connect them right now to observable predictions, there's no use talking about them. It's part of a much bigger picture—we have to understand how quantum mechanics and gravity play together long before we can ever hope to say definitively what the right answers are to these questions. SA

John Matson *is a former reporter and editor for* Scientific American *who has written extensively about astronomy and physics.*

MORE TO EXPLORE

The Fabric of the Cosmos. Brian Greene. Vintage, 2005.
The Cosmic Origins of Time's Arrow. Sean M. Carroll in *Scientific American,* Vol. 298, No. 6; pages 48–57; June 2008.
The Origin of the Universe and the Arrow of Time. Sean M. Carroll. CHAST 2009 Templeton Lecture. www.youtube.com/watch?v=G7WIJT--yso
From Eternity to Here: The Quest for the Ultimate Theory of Time. Sean M. Carroll. Plume, 2010.
From Eternity to Here: Sean M. Carroll's Quest to Understand Time. *Scientific American* Science Talk podcast. Published online March 30, 2010. www.scientificamerican.com/podcast/episode/from-eternity-to-here-sean-m-carrol-10-03-30
Sean M. Carroll's Cosmology Primer: www.preposterousuniverse.com/writings/preposterous

© 2018 Scientific American



© 2018 Scientific American

Lee Smolin is a senior, and founding, faculty member at the Perimeter Institute for Theoretical Physics in Waterloo, Ontario, and adjunct professor of physics at the University of Waterloo. In addition to his work on quantum gravity, he is interested in elementary particle physics, cosmology and the foundation of quantum theory. He is author of four books that probe philosophical issues raised by research in physics and cosmology, including *Time Reborn* (Knopf Canada, 2013).

PHYSICS

ATOMS OF SPACE AND TIME

We perceive space and time to be continuous, but if the amazing theory of loop quantum gravity is correct, they actually come in discrete pieces

By Lee Smolin

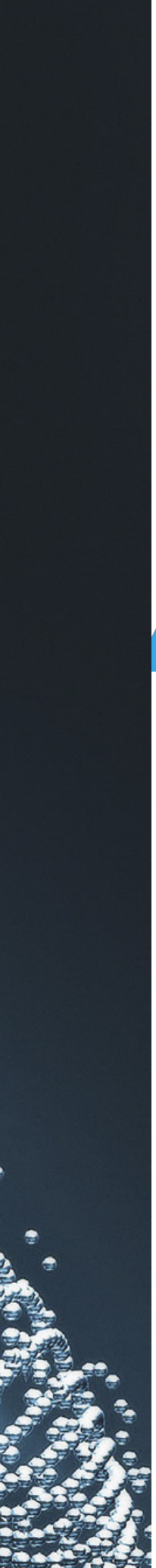


Illustration by Kenn Brown, Mondolithic Studios

© 2018 Scientific American

A LITTLE MORE THAN 100 YEARS AGO MOST PEOPLE—AND MOST SCIENTISTS—thought of matter as continuous. Although since ancient times some philosophers and scientists had speculated that if matter were broken up into small enough bits, it might turn out to be made up of very tiny atoms, few thought the existence of atoms could ever be proved. Today we have imaged individual atoms and have studied the particles that compose them. The granularity of matter is old news.

In recent decades physicists and mathematicians have asked if space is also made of discrete pieces. Is it continuous, as we learn in school, or is it more like a piece of cloth, woven out of individual fibers? If we could probe to size scales that were small enough, would we see "atoms" of space, irreducible pieces of volume that cannot be broken into anything smaller? And what about time: Does nature change continuously, or does the world evolve in a series of very tiny steps, acting more like a digital computer?

The past 30 years have seen great progress on these questions. A theory with the strange name of "loop quantum gravity" predicts that space and time are indeed made of discrete pieces. The picture revealed by calculations carried out within the framework of this theory is both simple and beautiful. The theory has deepened our understanding of puzzling phenomena having to do with black holes and the big bang. Best of all, it is possible that current experiments might be able to detect signals of the atomic structure of spacetime—if this structure actually exists—in the near future.

QUANTA

MY COLLEAGUES AND I developed the theory of loop quantum gravity while struggling with a long-standing problem in physics: Is it possible to develop a quantum theory of gravity? To explain why this is an important question—and what it has to do with the granularity of space and time—I must first say a bit about quantum theory and the theory of gravity.

The theory of quantum mechanics was formulated in the first quarter of the 20th century, a development that was closely connected with the confirmation that matter is made of atoms. The equations of quantum mechanics require that certain quantities, such as the energy of an atom, can come only in specific, discrete units. Quantum theory successfully predicts the properties and behavior of atoms and the elementary particles and forces that compose them. No theory in the history of science has been more successful than quantum theory. It underlies our understanding of chemistry, atomic and subatomic physics, electronics and even biology.

In the same decades that quantum mechanics was being

IN BRIEF

To understand the structure of space on the very smallest size scale, we must turn to a quantum theory of gravity. Einstein's general theory of relativity reveals that gravity is caused by the warping of space and time.

By carefully combining the fundamental principles of quantum mechanics and general relativity, physicists are led to the theory of "loop quantum gravity." In this theory, the allowed quantum states of space are represented by diagrams of lines and nodes called spin networks. Quantum spacetime corresponds to similar diagrams called spin foams, in which spin networks evolve over time.

Loop quantum gravity predicts that space comes in discrete lumps, the smallest of which is about a cubic Planck length, or 10^{-99} cubic centimeter. Time proceeds in discrete ticks of about a Planck time, or 10^{-43} second.

© 2018 Scientific American

formulated, Albert Einstein constructed his general theory of relativity, which is a theory of gravity. In his theory, the gravitational force arises as a consequence of space and time (which together form "spacetime") being curved by the presence of matter. A loose analogy is that of a bowling ball placed on a rubber sheet along with a marble that is rolling around nearby. The balls could represent the sun and the earth, and the sheet is space. The bowling ball creates a deep indentation in the rubber sheet, and the slope of this indentation causes the marble to be deflected toward the larger ball, as if some force—gravity—were pulling it in that direction. Similarly, any piece of matter or concentration of energy distorts the geometry of spacetime, causing other particles and light rays to be deflected toward it, a phenomenon we call gravity.

Quantum theory and Einstein's general theory of relativity have each separately been fantastically well confirmed by experiment—but no experiment has explored the regime where both theories predict significant effects. The problem is that quantum effects are most prominent at small size scales, whereas general relativistic effects require large masses, so it takes extraordinary circumstances to combine both conditions.

Allied with this hole in the experimental data is a huge conceptual problem: Einstein's general theory of relativity is thoroughly classical, or nonquantum. For physics as a whole to be logically consistent, there has to be a theory that somehow unites quantum mechanics and general relativity. This long-sought-after theory is called quantum gravity. Because general relativity deals in the geometry of spacetime, a quantum theory of gravity will in addition be a quantum theory of spacetime.

Physicists have developed a considerable collection of mathematical procedures for turning a classical theory into a quantum one. Many theoretical physicists and mathematicians have worked on applying those standard techniques to general relativity. Early results were discouraging. Calculations carried out in the 1960s and 1970s seemed to show that quantum theory and general relativity could not be successfully combined. Consequently, something fundamentally new seemed to be required, such as additional postulates or principles not included in quantum theory and general relativity, or new particles or fields, or new entities of some kind. Perhaps with the right additions or a new mathematical structure, a quantumlike theory could be developed that would successfully approximate general relativity in the nonquantum regime. To avoid spoiling the successful predictions of quantum theory and general relativity, the exotica contained in the full theory would remain hidden from experiment except in the extraordinary circumstances where both quantum theory and general relativity are expected to have large effects. Many different approaches along these lines have been tried, with names such as twistor theory, supergravity and string theory. After many years of study, however, none of these ideas has led to unambiguous predictions that can be tested experimentally. Thus, many physicists are reconsidering whether quantum theory and general relativity might be compatible after all.

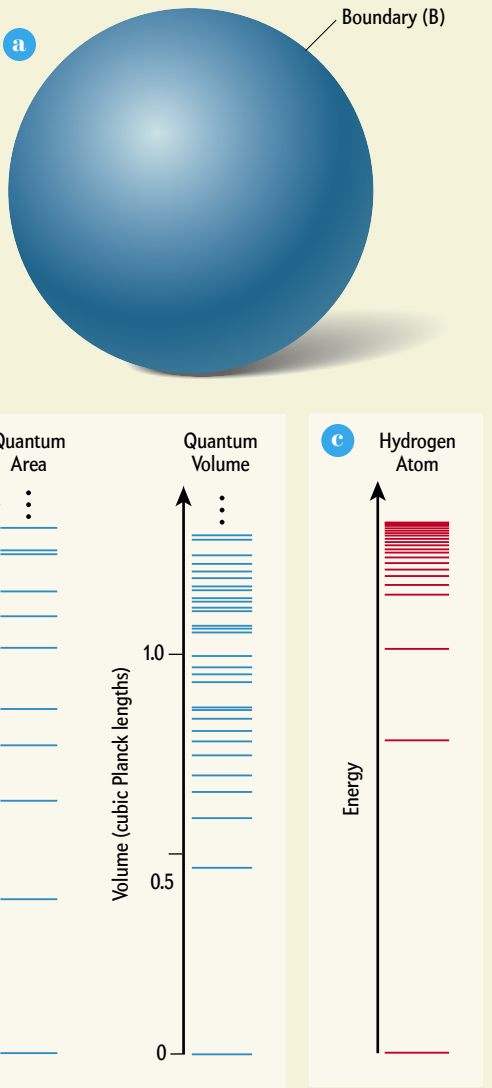
A BIG LOOPHOLE

IN THE MID-1980S a few of us—including Abhay Ashtekar, now at Pennsylvania State University, Ted Jacobson of the University of Maryland and Carlo Rovelli, now at Aix-Marseille

SIZE LIMITS

Quantum States of Volume and Area

A central prediction of the loop quantum gravity theory relates to volumes and areas. Consider a spherical shell that defines the boundary, B, of a region of space having some volume (*a*). According to classical (nonquantum) physics, the volume could be any positive real number. The loop quantum gravity theory says, however, that there is a nonzero absolute minimum volume (about one cubic Planck length, or 10^{-99} cubic centimeter), and it restricts the set of larger volumes to a discrete series of numbers. Similarly, according to the loop quantum gravity theory, there is a nonzero minimum area (about one square Planck length, or 10^{-66} square centimeter) and a discrete series of larger allowed areas. The discrete spectrum of allowed quantum areas and volumes (*b*) is broadly similar to the discrete quantum energy levels of a hydrogen atom (*c*).



NADIA STRASSER (*illustrations*); SOURCE: AREA AND VOLUME DATA FROM ROBERTO DE PIETRI AND CARLO ROVELLI *http://arxiv.org/abs/gr-qc/9602023*

© 2018 Scientific American

University in France—decided to reexamine the question of whether quantum mechanics could be combined consistently with general relativity using the standard techniques. We knew that the negative results from the 1970s had an important loophole. Those calculations assumed that the geometry of space is continuous and smooth, no matter how minutely we examine it, just as people had expected matter to be before the discovery of atoms. Some of our teachers and mentors had pointed out that if this assumption was wrong, the old calculations would not be reliable.

So we began searching for a way to do calculations without assuming that space is smooth and continuous. We insisted on not making any assumptions beyond the experimentally well tested principles of general relativity and quantum theory. In particular, we kept two key principles of general relativity at the heart of our calculations.

The first is known as background independence. This principle says that the geometry of spacetime is not fixed. Instead the geometry is an evolving, dynamical quantity. To find the geometry, one has to solve certain equations that include all the effects of matter and energy.

The second principle, known by the imposing name diffeomorphism invariance, is closely related to background independence. This principle implies that, unlike theories prior to general relativity, one is free to choose any set of coordinates to map spacetime and express the equations. A point in spacetime is defined only by what physically happens at it, not by its location according to some special set of coordinates (no coordinates are special). Diffeomorphism invariance is very powerful and is of fundamental importance in general relativity.

By carefully combining these two principles with the standard techniques of quantum mechanics, we developed a mathematical language that allowed us to do a computation to determine whether space is continuous or discrete. That calculation revealed, to our delight, that space is quantized. We had laid the foundations of our theory of loop quantum gravity. The term "loop," by the way, arises from how some computations in the theory involve small loops marked out in spacetime.

The calculations have been redone by a number of physicists and mathematicians using a range of methods. Over the years since, the study of loop quantum gravity has grown into a healthy field of research, with many contributors around the world; our combined efforts give us confidence in the picture of spacetime I will describe.

Ours is a quantum theory of the structure of spacetime at the smallest size scales, so to explain how the theory works we need to consider what it predicts for a small region or volume. In dealing with quantum physics, it is essential to specify precisely what physical quantities are to be measured. To do so, we consider a region somewhere that is marked out by a boundary, B [*see box on preceding page*]. The boundary may be defined by some matter, such as a cast-iron shell, or it may be defined by the geometry of spacetime itself, as in the event horizon of a black hole (a surface from within which even light cannot escape the black hole's gravitational clutches).

What happens if we measure the volume of the region? What are the possible outcomes allowed by both quantum theory and diffeomorphism invariance? If the geometry of space is continuous, the region could be of any size and the measurement result could be any positive real number; in particular, it could be as close as one wants to zero volume. But if the geometry is granular, then the measurement result can come from just a discrete set of numbers and it cannot be smaller than a certain minimum possible volume. The question is similar to asking how much energy electrons orbiting an atomic nucleus have. Classical mechanics predicts that an electron can possess any amount of energy, but quantum mechanics allows only specific energies (amounts in between those values do not occur). The difference is like that between the measure of something that flows continuously, like the 19th-century conception of water, and something that can be counted, like the atoms in that water.

The theory of loop quantum gravity predicts that space is like atoms: there is a discrete set of numbers that the volume-measuring experiment can return. Volume comes in distinct pieces. Another quantity we can measure is the area of the boundary B. Again, calculations using the theory return an unambiguous result: the area of the surface is discrete as well. In other words, space is not continuous. It comes only in specific quantum units of area and volume.

The possible values of volume and area are measured in units of a quantity called the Planck length. This length is related to the strength of gravity, the size of quanta and the speed of light. It measures the scale at which the geometry of space is no longer continuous. The Planck length is very small: 10^{-33} centimeter. The smallest possible nonzero area is about a square Planck length, or 10^{-66} cm^2. The smallest nonzero volume is approximately a cubic Planck length, 10^{-99} cm^3. Thus, the theory predicts that there are about 10^{99} atoms of volume in every cubic centimeter of space. The quantum of volume is so tiny that there are more such quanta in a cubic centimeter than there are cubic centimeters in the visible universe (10^{85}).

SPIN NETWORKS

WHAT ELSE does our theory tell us about spacetime? To start with, what do these quantum states of volume and area look like? Is space made up of a lot of little cubes or spheres? The answer is no—it's not that simple. Nevertheless, we can draw diagrams that represent the quantum states of volume and area. To those of us working in this field, these diagrams are beautiful because of their connection to an elegant branch of mathematics.

To see how these diagrams work, imagine that we have a lump of space shaped like a cube [*see box on opposite page*]. In our diagrams, we would depict this cube as a dot, which represents the volume, with six lines sticking out, each of which represents one of the cube's faces. We have to write a number next to the dot to specify the quantity of volume, and on each line we write a number to specify the area of the face that the line represents.

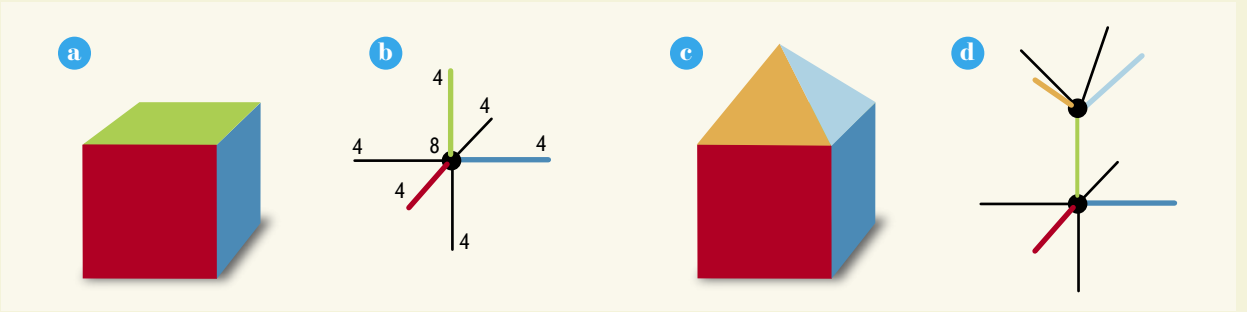
Next, suppose we put a pyramid on top of the cube. These two polyhedra, which share a common face, would be depicted as two dots (two volumes) connected by one of the lines (the face that joins the two volumes). The cube has five other faces (five lines sticking out), and the pyramid has four (four lines sticking out). It is clear how more complicated arrangements involving polyhedra other than cubes and pyramids could be depicted with these dot-and-line diagrams: each polyhedron of volume becomes a dot, or node, and each flat face of a polyhedron becomes a line, and the lines join the nodes in the way

© 2018 Scientific American

Visualizing Quantum States of Volume

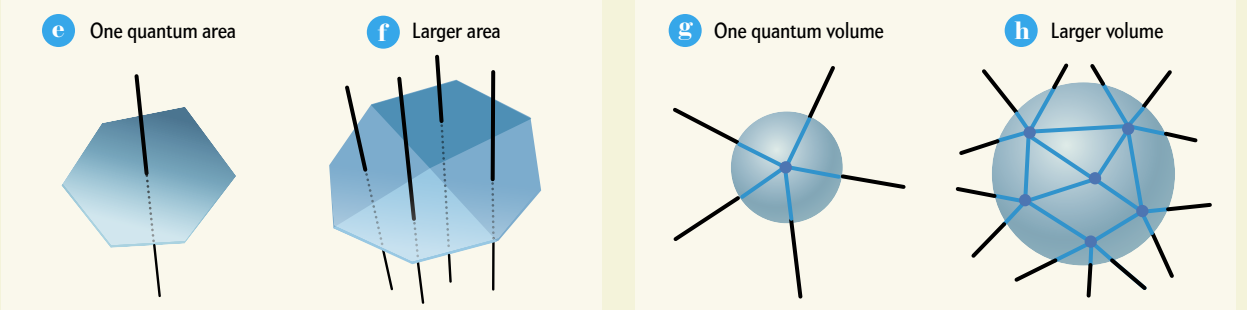
Diagrams called spin networks are used by physicists who study loop quantum gravity to represent quantum states of space at a minuscule scale. Some such diagrams correspond to polyhedra-shaped volumes. For example, a cube (*a*) consists of a volume enclosed within six square faces. The corresponding spin network (*b*) has a dot, or node, representing the volume and six lines that represent the six faces. The complete spin network has a number at the node to indicate the cube's volume and a number on each line to indicate the area of the corresponding face. Here the volume is eight cubic Planck lengths, and the faces are each four square Planck lengths. (The rules of loop quantum gravity restrict the allowed volumes and areas to specific quantities: only certain combinations of numbers are allowed on the lines and nodes.)

If a pyramid sat on the cube's top face (*c*), the line representing that face in the spin network would connect the cube's node to the pyramid's node (*d*). The lines corresponding to the four exposed faces of the pyramid and the five exposed faces of the cube would stick out from their respective nodes. (The numbers have been omitted for simplicity.)



In general, in a spin network, one quantum of area is represented by a single line (*e*), whereas an area composed of many quanta is represented by many lines (*f*). Similarly, a quantum of volume is represented by one node (*g*), whereas a larger volume takes many nodes (*h*). If we have a region of space defined by a spherical shell, the volume inside the shell is given by the sum of all the enclosed nodes and its surface area is given by the sum of all the lines that pierce it.

The spin networks are more fundamental than the polyhedra: any arrangement of polyhedra can be represented by a spin network in this fashion, but some valid spin networks represent combinations of volumes and areas that cannot be drawn as polyhedra. Such spin networks would occur when space is curved by a strong gravitational field or in the course of quantum fluctuations of the geometry of space at the Planck scale.



that the faces join the polyhedra together. Mathematicians call these line diagrams graphs.

Now, in our theory, we throw away the drawings of polyhedra and just keep the graphs. The mathematics that describes the quantum states of volume and area gives us a set of rules for how the nodes and lines can be connected and what numbers can go where in a diagram. Every quantum state corresponds to one of these graphs, and every graph that obeys the rules corresponds to a quantum state. The graphs are a convenient shorthand for all the possible quantum states of space. (The mathematics and other details of the quantum states are too complicated to discuss here; the best we can do is show some of the related diagrams.)

The graphs are a better representation of the quantum states than the polyhedra are. In particular, some graphs connect in strange ways that cannot be converted into a tidy picture of polyhedra. For example, whenever space is curved, the polyhedra will not fit together properly in any drawing we could do, yet we can still draw a graph. Indeed, we can take a

© 2018 Scientific American

graph and from it calculate how much space is distorted. Because the distortion of space is what produces gravity, this is how the diagrams form a quantum theory of gravity.

For simplicity, we often draw the graphs in two dimensions, but it is better to imagine them filling three-dimensional space because that is what they represent. Yet there is a conceptual trap here: the lines and nodes of a graph do not live at specific locations in space. Each graph is defined only by the way its pieces connect together and how they relate to well-defined boundaries such as boundary B. The continuous, three-dimensional space that you are imagining the graphs occupy *does not exist* as a separate entity. All that exist are the lines and nodes; they *are* space, and the way they connect defines the geometry of space.

These graphs are called spin networks because the numbers on them are related to quantities called spins. Roger Penrose of the University of Oxford first proposed in the early 1970s that spin networks might play a role in theories of quantum gravity. We were very pleased when we found, in 1994, that precise calculations confirmed his intuition. Readers familiar with Feynman diagrams should note that our spin networks are not Feynman diagrams, despite the superficial resemblance. Feynman diagrams represent quantum interactions between particles, which proceed from one quantum state to another. Our diagrams represent fixed quantum states of spatial volumes and areas.

The individual nodes and edges of the diagrams represent extremely small regions of space: a node is typically a volume of about one cubic Planck length, and a line is typically an area of about one square Planck length. But in principle, nothing limits how big and complicated a spin network can be. If we could draw a detailed picture of the quantum state of our universe—the geometry of its space, as curved and warped by the gravitation of galaxies and black holes and everything else—it would be a gargantuan spin network of unimaginable complexity, with approximately 10^{184} nodes.

These spin networks describe the geometry of space. But what about all the matter and energy contained in that space? How do we represent particles and fields occupying positions and regions of space? Particles, such as electrons, correspond to certain types of nodes, which are represented by adding more labels on nodes. Fields, such as the electromagnetic field, are represented by additional labels on the lines of the graph. We represent particles and fields moving through space by these labels moving in discrete steps on the graphs.

MOVES AND FOAMS

PARTICLES AND FIELDS are not the only things that move around. According to general relativity, the geometry of space changes in time. The bends and curves of space change as matter and energy move, and waves can pass through it like ripples on a lake. In loop quantum gravity, these processes are represented by changes in the graphs. They evolve in time by a succession of certain "moves" in which the connectivity of the graphs changes [*see box on opposite page*].

When physicists describe phenomena quantum-mechanically, they compute probabilities for different processes. We do the same when we apply loop quantum gravity theory to describe phenomena, whether it be particles and fields moving on the spin networks or the geometry of space itself evolving in time. Work by many people over the past 20 years has revealed an elegant set of rules for computing the probabilities of the different moves by which spin networks change in time. These rules express the quantum version of Einstein's equations of general relativity. With them, the theory is completely specified: we have a well-defined procedure for computing the probability of any process that can occur in a world that obeys the rules of our theory. It remains only to do the computations and work out predictions for what could be observed in experiments of one kind or another.

To discover the precise rules for computing probabilities, we had to follow Einstein in shifting our perspective from space to spacetime. Einstein's special and general theories of relativity join space and time together into the single, merged entity known as spacetime. The spin networks that represent space in loop quantum gravity theory accommodate the concept of spacetime by becoming what we call spin "foams." With the addition of another dimension—time—the lines of the spin networks grow to become two-dimensional surfaces and the nodes grow to become lines. Transitions where the spin networks change (the moves discussed earlier) are now represented by nodes where the lines meet in the foam. The spin-foam picture of spacetime was proposed by several people, including Rovelli, Mike Reisenberger (now at the University of Montevideo in Uruguay), John Barrett of the University of Nottingham in England, Louis Crane of Kansas State University, John Baez of the University of California, Riverside, and Fotini Markopoulou, then at the Perimeter Institute for Theoretical Physics in Ontario.

In the spacetime way of looking at things, a snapshot at a specific time is like a slice cutting across the spacetime. Taking such a slice through a spin foam produces a spin network. But it would be wrong to think of such a slice as moving continuously, like a smooth flow of time. Instead, just as space is defined by a spin network's discrete geometry, time is defined by the sequence of distinct moves that rearrange the network [*see box on opposite page*]. In this way, time also becomes discrete. Time flows not like a river but like the ticking of a clock, with "ticks" that are about as long as the Planck time: 10^{-43} second. Or, more precisely, time in our universe flows by the ticking of innumerable clocks—in a sense, at every location in the spin foam where a quantum "move" takes place, a clock at that location has ticked once.

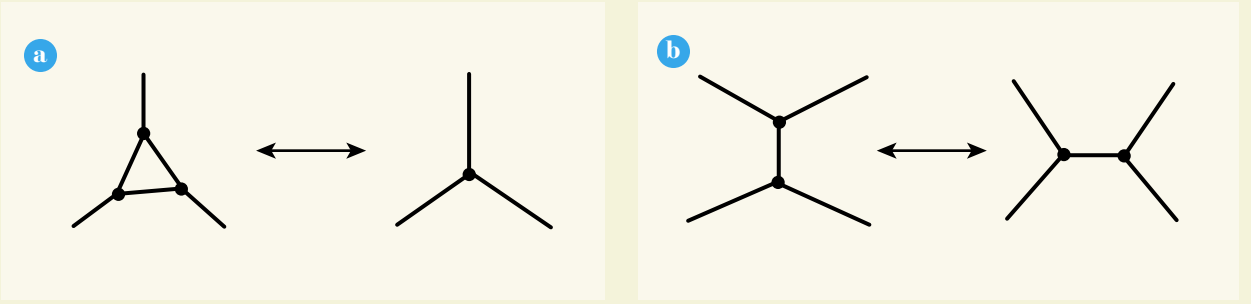
PREDICTIONS AND TESTS

I HAVE OUTLINED what loop quantum gravity has to say about space and time at the Planck scale, but we cannot verify the theory directly by examining spacetime on that scale. It is too small. So how can we test the theory? An important test is whether one can derive classical general relativity as an approximation to loop quantum gravity. In other words, if the spin networks are like the threads woven into a piece of cloth, this is analogous to asking whether we can compute the right elastic properties for a sheet of the material by averaging over thousands of threads. Similarly, when averaged over many Planck lengths, do spin networks describe the geometry of space and its evolution in a way that agrees roughly with the "smooth cloth" of Einstein's classical theory? This is a challenging problem, but with the right rules for computing probabilities, we have made a lot of progress. During the past few years the work of several groups has

© 2018 Scientific American

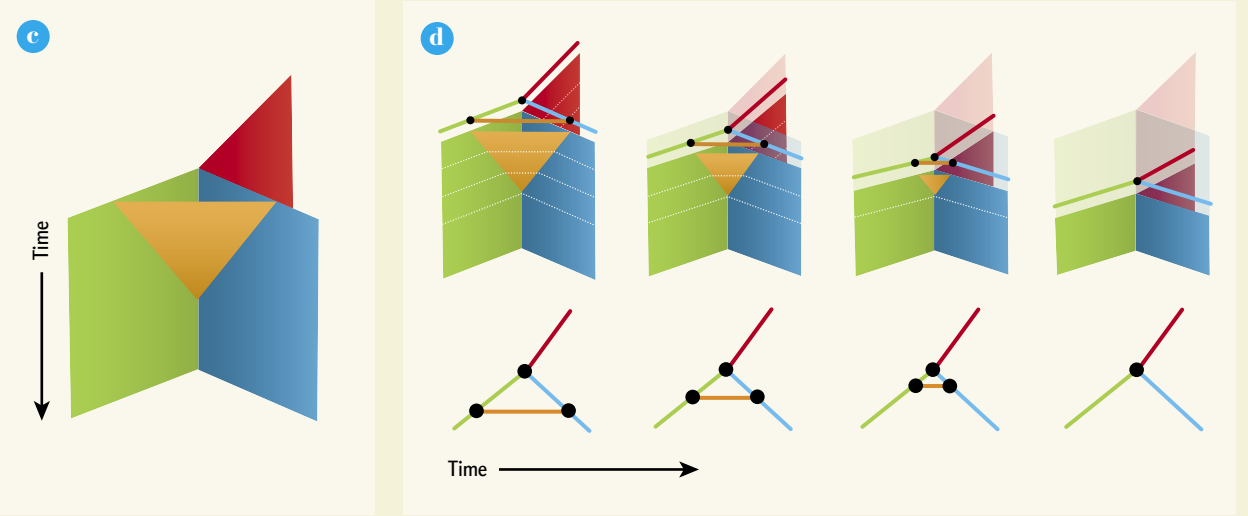
Evolution of Geometry in Time

Changes in the shape of space—such as those occurring when matter and energy move around within it and when gravitational waves flow by—are represented by discrete rearrangements, or moves, of the spin network. In *a,* a connected group of three volume quanta merge to become a single volume quantum; the reverse process can also occur. In *b,* two volumes divide up space and connect to adjoining volumes in a different way. Represented as polyhedra, the two polyhedra would merge on their common face and then split like a crystal cleaving on a different plane. These spin-network moves take place not only when large-scale changes in the geometry of space occur but also incessantly as quantum fluctuations at the Planck scale.



Another way to represent moves is to add the time dimension to a spin network—the result is called a spin foam (*c*). The lines of the spin network become planes, and the nodes become lines. Taking a slice through a spin foam at a particular time yields a spin network; taking a series of slices at different times produces frames of a movie showing the spin network evolving in time (*d*). But notice that the evolution, which at first glance appears to be smooth and continuous, is in fact discontinuous. All the spin networks that include the orange line (*first three frames shown*) represent exactly the same geometry of space. The length of the orange line does not matter—all that matters for the geometry is how the lines are connected and what number labels each line. Those are what define how the quanta of volume and area are arranged and how big they are. Thus, in *d,* the geometry remains constant during the first three frames, with three quanta of volume and six quanta of surface area. Then the geometry changes discontinuously, becoming a single quantum of volume and three quanta of surface area, as shown in the last frame. In this way, time as defined by a spin foam evolves by a series of abrupt, discrete moves, not by a continuous flow.

Although speaking of such sequences as frames of a movie is helpful for visualization, the more correct way to understand the evolution of the geometry is as discrete ticks of a clock. At one tick, the orange quantum of area is present; at the next tick, it is gone—in fact, the disappearance of the orange quantum of area defines the tick. The difference in time from one tick to the next is approximately the Planck time, 10^{-43} second. But time does not exist in between the ticks; there is no "in between," in the same way that there is no water in between two adjacent molecules of water.

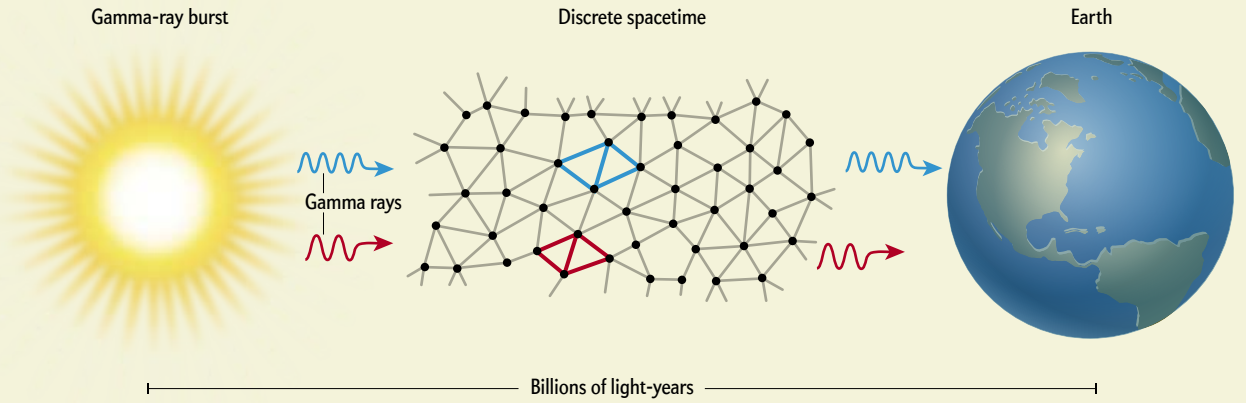


SOURCES: *A* AND *B* ADAPTED FROM FOTINI MARKOPOULOU *http://arxiv.org/abs/gr-qc/9704013*; C ADAPTED FROM CARLO ROVELLI *http://arxiv.org/abs/gr-qc/9806121*

© 2018 Scientific American

An Experimental Test

Radiation from distant cosmic explosions called gamma-ray bursts might provide a way to test whether the theory of loop quantum gravity is correct. Gamma-ray bursts occur billions of light-years away and emit a huge amount of gamma rays within a short span. According to loop quantum gravity, each photon occupies a region of lines at each instant as it moves through the spin network that is space (in reality, a very large number of lines, not just the five depicted here). The discrete nature of space may cause higher-energy gamma rays to travel slightly faster than lower-energy ones. The difference is tiny, but its effect steadily accumulates during the rays' billion-year voyage. If a burst's gamma rays arrive at Earth at slightly different times according to their energy, that would be evidence that quantum gravity implies a modification in special relativity. The Fermi satellite, launched in 2008, has the sensitivity required to discover this effect but has so far not detected any difference in arrival time between high-and lower-energy gamma rays.



made it possible to assert that when the elements of the discrete geometry are large, compared with the Planck scale, they behave in a way that is a good approximation of Einstein's equations of general relativity. In the same approximation, it has been shown that gravitons—the quanta associated with gravitational waves—travel and scatter from one another precisely as Einstein's general theory of relativity dictates.

Another fruitful test is to see what loop quantum gravity has to say about one of the long-standing mysteries of gravitational physics and quantum theory: the thermodynamics of black holes—in particular, their entropy, which is related to disorder. Physicists have computed predictions regarding black hole thermodynamics using a hybrid, approximate theory in which matter is treated quantum-mechanically but spacetime is not. A full quantum theory of gravity, such as loop quantum gravity, should be able to reproduce these predictions. Specifically, in the 1970s Jacob D. Bekenstein, now at the Hebrew University of Jerusalem, inferred that black holes must be ascribed an entropy proportional to their surface area. Shortly after, Stephen Hawking of the University of Cambridge deduced that black holes, particularly small ones, must emit radiation. These predictions are among the greatest results of theoretical physics in the past 40 years.

To do the calculation in loop quantum gravity, we pick the boundary B to be the event horizon of a black hole. When we analyze the entropy of the relevant quantum states, we get *precisely* the prediction of Bekenstein. Similarly, the theory reproduces Hawking's prediction of black hole radiation. In fact, it makes further predictions for the fine structure of Hawking radiation. If a microscopic black hole were ever observed, this prediction could be tested by studying the spectrum of radiation it emits. That may be far off in time, however, because we have no technology to make black holes, small or otherwise.

Indeed, an experimental test of any quantum theory of gravity would appear at first to be an immense technological challenge. The problem is that the characteristic effects described by the theory become significant only at the Planck scale, the very tiny size of the quanta of area and volume. The Planck scale is 16 orders of magnitude below the scale probed in the highest-energy particle accelerators currently planned (higher energy is needed to probe shorter-distance scales). Because we cannot reach the Planck scale with an accelerator, many people have held out little hope for the confirmation of quantum gravity theories.

A few imaginative young researchers in the past 20 years have thought up new ways to test the predictions of loop quantum gravity that can be done now. These methods depend on the propagation of light across the universe. When light moves through a medium, its wavelength suffers some distortions, leading to effects such as bending in water and the separation of different wavelengths, or colors. These effects may also occur for light and particles moving through the discrete space described by a spin network.

Unfortunately, the magnitude of the effects is proportional to the ratio of the Planck length to the wavelength. For visible light,

© 2018 Scientific American

this ratio is smaller than 10^{-28}; even for the most powerful cosmic rays ever observed, it is about one billionth. For any radiation we can observe, the effects of the granular structure of space are very small. What the young researchers spotted is that these effects accumulate when light travels a long distance. And we detect light and particles that come from billions of light-years away, from events such as gamma-ray bursts.

Gamma-ray bursts contain photons of very different energies. According to Einstein's special theory of relativity, all photons travel at the universal speed of light. Consequently, the photons emitted by a burst should arrive in the order they are emitted. But a possible consequence of the light traveling through a discrete spacetime is to modify this law so that a photon's speed depends very slightly on its energy. This implies that the quantum structure of spacetime comes into conflict with Einstein's special theory of relativity. This conflict is very small on usual scales but becomes detectable when we study light that has traveled for enormous distances, over which time the effect is amplified so that more energetic photons tend to arrive earlier than their less energetic siblings.

There are actually two different possibilities. The first is that quantum spacetime violates the basic relativity principle that velocity and rest are relative concepts. This would mean that there is an observer for whom the atoms of spacetime appear to be at rest, just like the atoms of a solid body.

The second is that the relativity principle is preserved but the laws of special relativity are modified in such a way that the time it takes photons to travel from the burst to the detector depends on their energy. This second possibility is called doubly special relativity; recently it has been subsumed in a deeper concept called relative locality.

Several experiments currently have the sensitivity required to sort out the fate of Einstein's special theory of relativity in the presence of quantum spacetime. The most important of these has been the Fermi Gamma-Ray Space Telescope, in orbit since June 2008. Measurements it has already made limit the scope of such an effect to just about the quantum gravity scale. Other observations—of the behavior of polarized radio waves from galaxies and of the behavior of very energetic cosmic rays—appear to confirm the validity of the relativity principle even at scales of quantum geometry. Observations by Fermi could either rule out or confirm the possibility that special relativity is modified by quantum spacetime.

Does loop quantum gravity make predictions for these experiments? The short answer is not yet. In the 1990s several calculations pointed to a violation of the relativity principle, but these turned out to be based on incorrect forms for the laws of spin-network evolution. The laws now understood to be correct do not violate the relativity principle. Whether they lead to a modification of the laws of special relativity is under investigation.

THE COSMOS

LOOP QUANTUM GRAVITY has opened up a new window through which we can study deep cosmological questions such as those relating to the origins of our universe. We can use the theory to study the earliest moments of time just after the big bang. General relativity predicts that there was a first moment of time, but this conclusion ignores quantum physics (because general relativity is not a quantum theory). Models of the very early universe based on loop quantum gravity, however, indicate that the big bang is actually a big bounce; before the bounce the universe was rapidly contracting. Theorists are now hard at work developing predictions for the early universe that may be testable in future cosmological observations. It is not impossible that in our lifetime we could see evidence of the time before the big bang.

Another possible observational signature of loop quantum gravity would be a breaking of the symmetry between left and right, which could be detected in observations of polarization in the cosmic background radiation. If this effect is present, then the universe seen in a mirror would be distinguishable from the universe seen directly. As pointed out by João Magueijo of Imperial College London and his collaborators, this is a natural consequence allowed by loop quantum gravity, and it could be detected in observations by the Planck satellite and other observatories.

Recent research on loop quantum gravity also addresses the unification of gravity with the other forces found in nature. It is even possible to incorporate extra dimensions and supersymmetry in the theory, if needed. But as is the case with string theory, no principle has emerged that would predict a unique unification of gravity with particle physics.

Many open questions remain to be answered in loop quantum gravity. While there is now good evidence that general relativity emerges as an approximation to the theory within certain limits, we would like to better understand how robust is this key property. And we must go beyond it to know what modifications of relativity theory are implied, as these could lead to observable effects.

In addition to loop quantum gravity, there are a few other background, independent approaches to quantum gravity that have yielded interesting results. These include causal dynamical triangulations, causal sets, quantum "graphity," matrix models and shape dynamics. Loop quantum gravity occupies a very important place in the development of physics. It is arguably *the* quantum theory of general relativity because it makes no extra assumptions beyond the basic principles of quantum theory and relativity theory. The remarkable departure that it makes—proposing a discontinuous spacetime described by spin networks and spin foams—emerges from the mathematics of the theory itself, rather than being inserted as an ad hoc postulate.

Still, everything I have discussed is theoretical. It could be that in spite of all I have described here, space really is continuous, no matter how small the scale we probe. Because this is science, in the end experiment will decide. The good news is that the decision may come soon. SA

MORE TO EXPLORE

The Life of the Cosmos. Lee Smolin. Oxford University Press, 1999.
The Quantum of Area? John Baez in *Nature*, Vol. 421, pages 702–703; February 2003.
How Far Are We from the Quantum Theory of Gravity? Lee Smolin. March 2003. http://arxiv.org/abs/hep-th/0303185
Welcome to Quantum Gravity. Special section. *Physics World*, Vol. 16, No. 11, pages 27–50; November 2003.
Three Roads to Quantum Gravity. Third edition. Lee Smolin. Basic Books, 2017.
Loop Quantum Gravity. Lee Smolin. Online at http://edge.org/memberbio/lee_smolin
The Order of Time. Carlo Rovelli. Riverhead Books. 2018.

scientificamerican.com/magazine/sa

© 2018 Scientific American

George Musser is a contributing editor for *Scientific American*, author of *The Complete Idiot's Guide to String Theory* (Alpha, 2008) and of *Spooky Action at a Distance* (Farrar, Straus and Giroux, 2015). This article won the 2011 Science Communication Award from the American Institute of Physics.

COSMOLOGY

Could Time End?

Yes. And no. For time to end seems both impossible and inevitable. Recent work in physics suggests a resolution to the paradox

By George Musser

IN BRIEF

Einstein's general theory of relativity predicts that time ends at moments called singularities, such as when matter reaches the center of a black hole or the universe collapses in a "big crunch." Yet singularities require matter to achieve conditions that are in fact physically impossible.

A way to resolve this paradox is to consider time's death as gradual rather than abrupt. Time might lose its many attributes one by one: its directionality, its notion of duration and its role in ordering events causally. Finally, time might give way to deeper, timeless physics.

IN OUR EXPERIENCE, NOTHING EVER REALLY ENDS. When we die, our bodies decay and the material in them returns to the earth and the air, allowing for the creation of new life. We live on in what comes after. But will that always be the case? Might there come a point sometime in the future when there is no "after"? Depressingly, modern physics suggests the answer is yes. Time itself could end. All activity would cease, and there would be no renewal or recovery. The end of time would be the end of endings.

This grisly prospect was an unanticipated prediction of Einstein's general theory of relativity, which provides our modern understanding of gravity. Before that theory, most physicists and philosophers thought time was a universal drumbeat, a steady rhythm that the cosmos marches to, never varying, wavering or stopping. Albert Einstein showed that the universe is more like a big polyrhythmic jam session. Time can slow down, or stretch out, or let it rip. When we feel the force of gravity, we are feeling time's rhythmic improvisation; falling objects are drawn to places where time passes more slowly. Time not only affects what matter does but also responds to what matter is doing, like drummers and dancers firing one

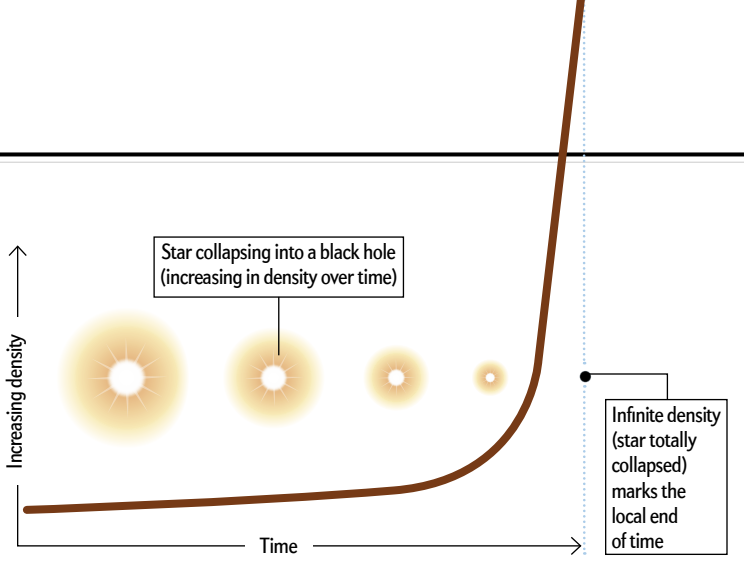
HOLLY LINDEM (clocks)

© 2018 Scientific American

© 2018 Scientific American

Ultimate Doomsday

Time can end in a disquieting variety of ways, according to Einstein's general theory of relativity. For instance, when a black hole forms, the density of matter increases, which intensifies the force of gravity, which further increases the density, which further intensifies gravity, and so it goes until density and gravity both become infinite—a condition known as a singularity, as depicted in the graph. Matter ceases to be, and time runs out for that region of space. A similar fate could befall the entire universe, as spelled out in the six scenarios.


Star collapsing into a black hole (increasing in density over time)
Increasing density
Time
Infinite density (star totally collapsed) marks the local end of time

Big Trouble for Time

BIG CRUNCH

Cosmic expansion slows down as the gravity of matter holds it back. Eventually it stops and goes into reverse, culminating with collapse back to a singularity that marks time's end. Once thought likely, this fate now seems doubtful. Not only is matter too sparse to act as a brake, but some unseen form of energy—dark energy—appears to be stepping on the accelerator.

BIG WHIMPER

The universe expands forever, becoming ever emptier and gloomier. Astronomers now consider this fate the likeliest. Although time never ends, it becomes increasingly pointless. The universe suffers "heat death"—a state of equilibrium in which every process is quickly undone, so that time has no clear forward progression and might not even come in well-defined units.

BIG RIP

The universe tears itself to pieces. That can happen if dark energy is not constant, as most models suppose, but gains in power. Hypothesized in 1999, this extrapotent dark energy goes by the name of phantom energy. It drives the universe to expand by an infinite amount—even atoms get shredded—and finishes off time. In some scenarios, the end comes about 20 billion years from now.

BIG FREEZE

The universe fills up with phantom energy and reaches infinite density, while expanding by only a finite amount. Any surviving wisps of matter are locked in place, unable to move, and time seizes up. A localized big freeze can occur if our universe is a membrane that is moving in a higher-dimensional space (as postulated by string theory) and begins to whip around violently.

BIG BRAKE

Dark energy shifts from driving cosmic expansion to retarding it, bringing the growth of the universe to a screeching halt—the rate of deceleration is infinite. First conjectured in 2004, the event is traumatic, with cosmic structures subjected to tidal forces of infinite strength. Although other quantities remain finite, the situation has unhappy consequences for time.

BIG LURCH

Conceived of in 2004, a so-called sudden singularity does not require dark energy; all it takes is for ordinary matter to work itself up into a frenzy. Pressure forces go infinite, while the density and cosmic expansion rate remain in safe territory. Time might or might not carry on. Astronomical observations cannot rule out such a calamity as soon as nine million years from now.

another up into a rhythmic frenzy. When things get out of hand, though, time can go up in smoke like an overexcited drummer who spontaneously combusts.

The moments when that happens are known as singularities. The term actually refers to any boundary of time, be it beginning or end. The best known is the big bang, the instant 13.8 billion years ago when our universe—and, with it, time—burst into existence and began expanding. If the universe ever stops expanding and starts contracting again, it will go into something like the big bang in reverse—the big crunch—and bring time crashing to a halt.

Time needn't perish everywhere. Relativity says it expires inside black holes while carrying on in the universe at large. Black holes have a well-deserved reputation for destructiveness, but they are even worse than you might think. If you fell into one, you would not only be torn to shreds, but your remains would eventually hit a singularity at the center of the hole, and your time line would end. No new life would emerge from your ashes; your molecules would not get recycled. Like a character reaching the last page of a novel, you would not suffer mere death but existential apocalypse.

It took physicists decades to accept that relativity theory would predict something so unsettling as death without rebirth. To this day, they aren't quite sure what to make of it. Singularities are arguably the leading reason that physicists seek to create a unified theory of physics, which would merge Einstein's brainchild with quantum mechanics to create a quantum theory of gravity. They do so partly in the hope they might explain singularities away. Still, you need to be careful what you wish for. Time's end is hard to imagine, but time's not ending may be equally paradoxical.

Well before Einstein came along, philosophers through the ages had debated whether time could be mortal. Immanuel Kant considered the issue to be an "antinomy"—something you

LUCY READING-IKKANDA (*chart*); JASON LEE (*illustrations*)

© 2018 Scientific American

could argue both ways, leaving you not knowing what to think.

My father-in-law found himself on one horn of this dilemma when he showed up at an airport one evening only to find that his flight had long since departed. The people at the check-in counter chided him, saying he should have known that the scheduled departure time of "12 A.M." meant the first thing in the morning. Yet my father-in-law's confusion was understandable. Officially there is no such time as "12 A.M." Midnight is both the end of one day and start of the next. In 24-hour time notation, it is both 2400 and 0000.

Aristotle appealed to a similar principle when he argued that time can have neither beginning nor end. Every moment is both the end of an era and the start of something new; every event is both the outcome of something and the cause of something else. So how could time possibly end? What would prevent the last event in history from leading to another? Indeed, how would you even define the end of time when the very concept of "end" presupposes time? "It is not logically possible for time to have an end," asserts University of Oxford philosopher Richard Swinburne. But if time cannot end, then the universe must be infinitely long-lived, and all the riddles posed by the notion of infinity come rushing in. Philosophers have thought it absurd that infinity could be anything but a mathematical idealization.

The triumph of the big bang theory and the discovery of black holes seemed to settle the question. The universe is shot through with singularities and could suffer a distressing variety of temporal cataclysms; even if it evades the big crunch, it might get done in by the big rip, the big freeze or the big brake [*see box on opposite page*]. But then ask what singularities (big or otherwise) actually are, and the answer is no longer so clear. "The physics of singularities is up for grabs," says Lawrence Sklar of the University of Michigan, a leading philosopher of physics.

The very theory that begat these monsters suggests they cannot really exist. At the big bang singularity, for example, relativity theory says that the precursors of every single galaxy we see were squashed into a single mathematical point—not just a tiny pinprick but a true point of zero size. Likewise, in a black hole, every single particle of a hapless astronaut gets compacted into an infinitesimal point. In both cases, calculating the density means dividing by zero volume, yielding infinity. Other types of singularities do not involve infinite density but an infinite something else.

TIME FOR A NEW THEORY

ALTHOUGH MODERN PHYSICISTS do not feel quite the same aversion to infinity that Aristotle and Kant did, they still take it as a sign they have pushed a theory too far. For example, consider the standard theory of ray optics taught in middle school. It beautifully explains eyeglass prescriptions and funhouse mirrors. It also predicts that a lens focuses light from a distant source to a single mathematical point, producing a spot of infinite intensity. In reality, light gets focused not to a point but to a bull's-eye pattern. Its intensity may be high but is always finite. Ray optics errs because light is not really a ray but a wave.

In a similar vein, nearly all physicists presume that cosmic singularities actually have a finite, if high, density. Relativity theory errs because it fails to capture some important aspect of gravity

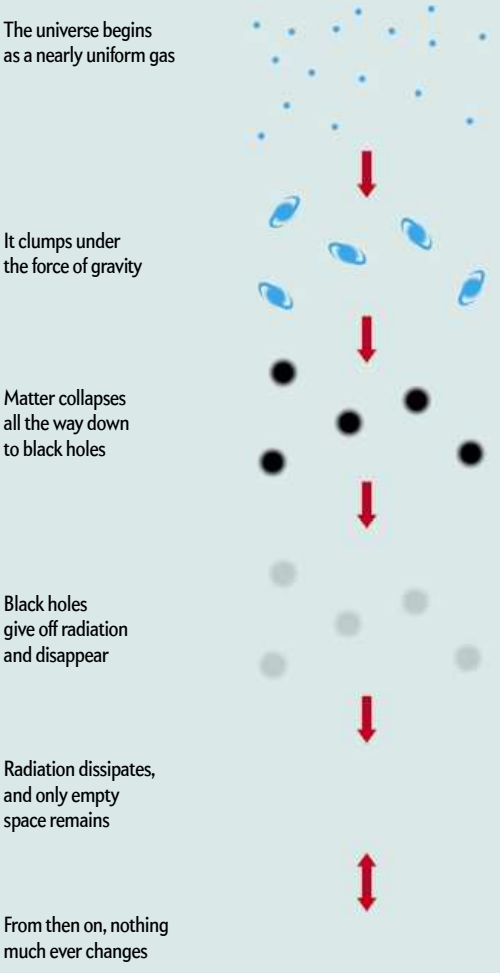
The Four Stages of the End of Time

The end of time might be a step-by-step process, as the universe regresses to a more primitive state in which time has no meaning. (The sequence shown is not rigid; the steps might overlap or occur in a different order.)

1. LOSS OF DIRECTIONALITY

Broken Arrow of Time

Time will stop marching forward when the universe exhausts its useful energy and reaches a condition of general stasis. The scenario shown occurs in an eternally expanding universe, but time can lose its directionality in other scenarios as well. From then on, the only activity will be random fluctuations of density and energy, causing clocks, if there are any left, merely to jiggle back and forth.

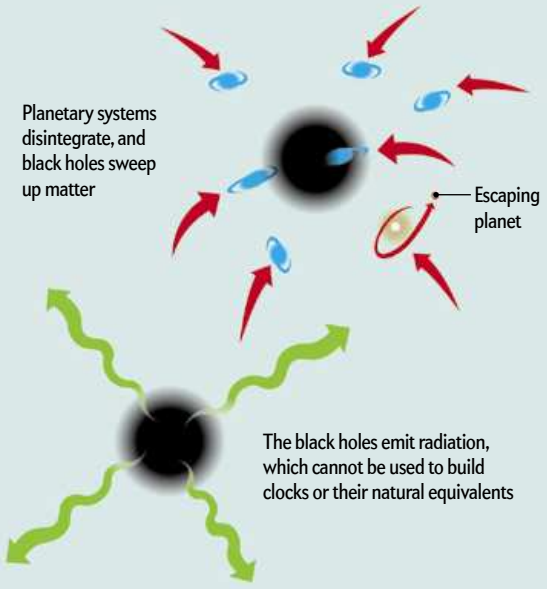

The universe begins as a nearly uniform gas
It clumps under the force of gravity
Matter collapses all the way down to black holes
Black holes give off radiation and disappear
Radiation dissipates, and only empty space remains
From then on, nothing much ever changes

© 2018 Scientific American

2. LOSS OF DURATION

Time Cannot Be Told

The concept of duration will become meaningless when all systems that mark out regular time intervals fall apart or get swallowed by black holes. Energy may leak back out of the black holes, but it does so as radiation—that is, as photons and other massless particles. Because such particles have no fixed scale and do not change with time, they cannot be used as the basis for new clocks.


Planetary systems disintegrate, and black holes sweep up matter
Escaping planet
The black holes emit radiation, which cannot be used to build clocks or their natural equivalents

or matter that comes into play near singularities and keeps the density under control. "Most people would say that they signal that the theory is breaking down there," says physicist James B. Hartle of the University of California, Santa Barbara.

To figure out what goes on will take a more encompassing theory, a quantum theory of gravity. Physicists are still working on such a theory, but they figure that it will incorporate the central insight of quantum mechanics: that matter, like light, has wavelike properties. These properties should smear the putative singularity into a small wad, rather than a point, and thereby banish the divide-by-zero error. If so, time may not, in fact, end.

Physicists argue it both ways. Some think time does end. The trouble with this option is that the known laws of physics operate within time and describe how things move and evolve. Time's end points are off the reservation; they would have to be governed not just by a new law of physics but by a new type of law of physics, one that eschews temporal concepts such as motion and change in favor of timeless ones such as geometric elegance. In one proposal several years ago Brett McInnes of the National University of Singapore drew on ideas from the leading candidate for a quantum theory of gravity—string theory. He suggested that the primordial wad of a universe had the shape of a torus; because of mathematical theorems concerning tori, it had to be perfectly uniform and smooth. At the big crunch or a black hole singularity, however, the universe could have any shape whatsoever, and the same mathematical reasoning need not apply; the universe would in general be extremely raggedy. Such a geometric law of physics differs from the usual dynamical laws in a crucial sense: it is not symmetrical in time. The end wouldn't just be the beginning played backward.

Other quantum gravity researchers think that time stretches on forever, with neither beginning nor end. In their view, the big bang was simply a dramatic transition in the eternal life of the universe. Perhaps the pre-bangian universe started to undergo a big crunch and turned around when the density got too high—a big bounce. Artifacts of this prehistory may even have made it through to the present day. By similar reasoning, the singular wad at the heart of a black hole would boil and burble like a miniaturized star. If you fell into a black hole, you would die a painful death, but at least your time line would not end. Your particles would plop into the wad and leave a distinct imprint on it, one that future generations might see in the feeble glow of light the hole gives off.

By supposing that time marches on, proponents of this approach avoid the need to speculate about a new type of law of physics. Yet they, too, run into trouble. For instance, the universe gets steadily more disordered with time. If it has been around forever, why is it not in total disarray by now? As for a black hole, how would the light bearing your imprint possibly manage to escape the hole's gravitational clutches?

The bottom line is that physicists struggle with antinomy no less than philosophers have. The late John Archibald Wheeler, a pioneer of quantum gravity, wrote, "Einstein's equation says 'this is the end' and physics says 'there is no end.'" Faced with this dilemma, some people throw up their hands and conclude that science can never resolve whether time ends. For them, the boundaries of time are also the boundaries of reason and empirical observation. Yet others think the puzzle just requires some fresh thinking. "It is not outside the scope of physics," says physicist Gary Horowitz of U.C. Santa Barbara. "Quantum gravity should be able to provide a definite answer."

HOW TIME SLIPS AWAY

THE HAL 9000 may have been a computer, but he was probably the most human character in *2001: A Space Odyssey*—expressive, resourceful, a bundle not just of wires but also of contradictions. Even his death was evocative of human death. It was not an event but a process. As Dave slowly pulled out his circuit boards, HAL lost his mental faculties one by one and described how it felt. He articulated the experience of regression in a way that people who die are often unable to. Human life is a complex feat of organization, the most complex known to science, and its emergence or submergence passes through the twilight between life and not life. Modern medicine shines a lantern into that twilight, as doctors save premature babies who once would have been lost and bring back people who have passed what was once a point of no return.

As physicists and philosophers struggle to grasp the end of time, many see parallels with the end of life. Just as life emerges out of lifeless molecules that organize themselves, time might emerge from some timeless stuff that brings itself to or-

© 2018 Scientific American

der [see "Is Time an Illusion?" by Craig Callender, on page 14]. A temporal world is a highly structured one. Time tells us when events occur, for how long and in what order. Perhaps this structure was not imposed from the outside but arose from within. What can be made can be unmade. When the structure crumbles, time ends.

By this thinking, time's demise is no more paradoxical than the disintegration of any other complex system. One by one, time loses its features and passes through the twilight from existence to nonexistence.

The first to go might be its unidirectionality—its "arrow" pointing from past to future. Physicists have recognized since the mid-19th century that the arrow is a property not of time per se but of matter. Time is inherently bidirectional; the arrow we perceive is simply the natural degeneration of matter from order to chaos, a syndrome that anyone who lives with pets or young children will recognize. (The original orderliness might owe itself to the geometric principles that McInnes conjectured.) If this trend keeps up, the universe will approach a state of equilibrium, or "heat death," in which it cannot possibly get any messier. Individual particles will continue to reshuffle themselves, but the universe as a whole will cease to change, any surviving clocks will jiggle in both directions and the future will become indistinguishable from the past. A few physicists have speculated that the arrow might reverse so that the universe could set about tidying itself up, but for mortal creatures whose very existence depends on a forward arrow of time, such a reversal would mark an end to time as surely as heat death would.

LOSING TRACK OF TIME

MORE RECENT RESEARCH suggests that the arrow is not the only feature that time might lose as it suffers death by attrition. Another could be the concept of duration. Time as we know it comes in amounts: seconds, days, years. If it didn't, we could tell that events occurred in chronological order but couldn't tell how long they lasted. That scenario is what University of Oxford physicist Roger Penrose presents in a book first published in the U.S. in 2011, *Cycles of Time: An Extraordinary New View of the Universe.*

Throughout his career, Penrose really seems to have had it in for time. He and University of Cambridge physicist Stephen Hawking showed in the 1960s that singularities do not arise only in special settings but should be everywhere. He has also argued that matter falling into a black hole has no afterlife and that time has no place in a truly fundamental theory of physics.

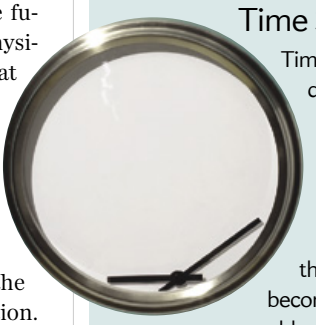
In his latest assault, Penrose begins with a basic observation about the very early universe. It was like a box of Legos that had just been dumped out on the floor and not yet assembled—a mishmash of quarks, electrons and other elementary particles. From them, structures such as atoms, molecules, stars and galaxies had to piece themselves together step by step. The first step was the creation of protons and neutrons, which consist of three quarks apiece and are about a femtometer (10^{-15} meter) across. They came together about 10 microseconds after the big bang (or big bounce, or whatever it was).

Before then, there were no structures at all—nothing was made up of pieces that were bound together. So there was nothing that could act as a clock. The oscillations of a clock rely on a well-defined reference such as the length of a pendulum, the distance between two mirrors or the size of atomic orbitals. No such reference yet existed. Clumps of particles might have come together temporarily, but they could not tell time, because they had no fixed size. Individual quarks and electrons could not serve as a reference, because they have no size, either. No matter how closely particle physicists zoom in on one, all they see is a point. The only sizelike attribute these particles have is their so-called Compton wavelength, which sets the scale of quantum effects and is inversely proportional to mass. And they lacked even this rudimentary scale prior to a time of about 10 picoseconds after the big bang, when the process that endowed them with mass had not yet occurred.

"There's no sort of clock," Penrose says. "Things don't know how to keep track of time." Without anything capable of marking out regular time intervals, either an attosecond or a femtosecond could pass, and it made no difference to particles in the primordial soup.

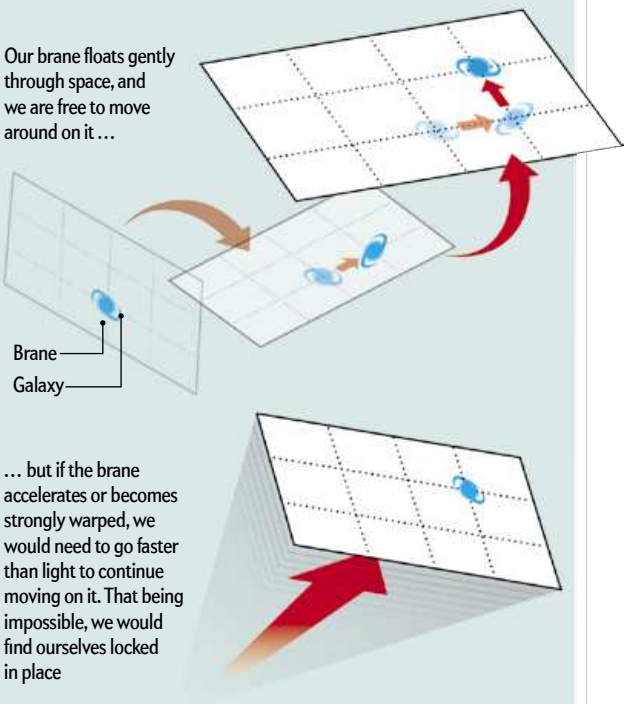
Penrose proposes that this situation describes not only the

3. LOSS OF CAUSALITY



Time Shades into Space

Time may be reduced to just another dimension of space, breaking the link between cause and effect. One way that can happen is if our universe is a "brane" floating through a higher-dimensional spacetime, and this brane begins to whip around so fast that the time dimension bends over and becomes a spatial one, producing what we would experience as a "big freeze."

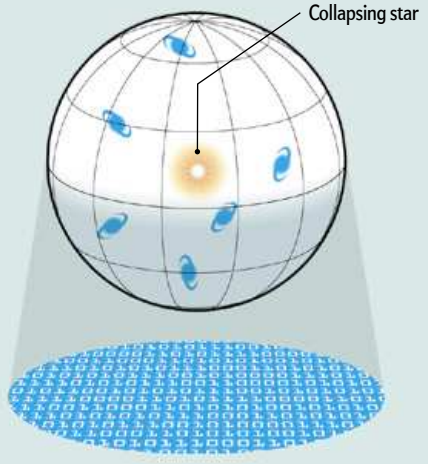

Our brane floats gently through space, and we are free to move around on it …
Brane
Galaxy
… but if the brane accelerates or becomes strongly warped, we would need to go faster than light to continue moving on it. That being impossible, we would find ourselves locked in place

© 2018 Scientific American

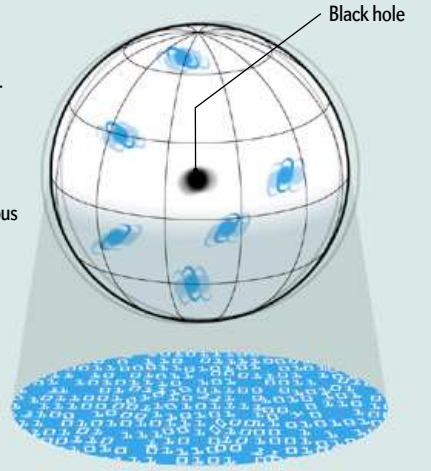
4. LOSS OF STRUCTURE

Geometry Dissolves

Time disappears altogether as the universe descends into anarchy. This anarchy breaks out at the deepest level of reality, even deeper than that of the known particles and forces. Processes become so complex that they cannot be said to occur at specific places and times. One way to grasp this idea is to consider the so-called holographic principle.



Our universe may really be 2-D, but certain regularities make it look 3-D—as if it were the projection of a hologram



Near black holes, the universe becomes increasingly chaotic, causing the locations and timing of events to become ambiguous

Eventually the 3-D projection breaks up altogether, and all that remains is a highly complex 2-D system

distant past but also the distant future. Long after all the stars wink out, the universe will be a grim stew of black holes and loose particles; then even the black holes will decay away and leave only the particles. Most of those particles will be massless ones such as photons, and again clocks will become impossible to build. In alternative futures where the universe gets snuffed out by, say, a big crunch, clocks don't fare too well, either.

You might suppose that duration will continue to make sense in the abstract, even if nothing could measure it. But researchers question whether a quantity that cannot be measured even in principle really exists. To them, the inability to build a clock is a sign that time itself has been stripped of one of its defining features. "If time is what is measured on a clock and there are no clocks, then there is no time," says philosopher of physics Henrik Zinkernagel of the University of Granada in Spain, who also has studied the disappearance of time in the early universe.

Despite its elegance, Penrose's scenario does have its weak points. Not all the particles in the far future will be massless; at least some electrons will survive, and you should be able to build a clock out of them. Penrose speculates that the electrons will somehow go on a diet and shed their mass, but he admits he is on shaky ground. "That's one of the more uncomfortable things about this theory," he says. Also, if the early universe had no sense of scale, how was it able to expand, thin out and cool down?

If Penrose is on to something, however, it has a remarkable implication. Although the densely packed early universe and ever emptying far future seem like polar opposites, they are equally bereft of clocks and other measures of scale. "The big bang is very similar to the remote future," Penrose says. He boldly surmises that they are actually the same stage of a grand cosmic cycle. When time ends, it will loop back around to a new big bang. Penrose, a man who has spent his career arguing that singularities mark the end of time, may have found a way to keep it going. The slayer of time has become its savior.

TIME STANDS STILL

EVEN IF DURATION becomes meaningless, and the femtoseconds and attoseconds blur into one another, time isn't dead quite yet. It still dictates that events unfold in a sequence of cause and effect that is the same for all observers. In this respect, time is different from space. Two events that are adjacent within time—when I type on my keyboard, letters appear on my screen—are inextricably linked. But two objects that are adjacent within space—a keyboard and a Post-it note—might have nothing to do with each other. Spatial relations simply do not have the same inevitability that temporal ones do.

Under certain conditions, however, time could lose even this basic ordering function and become just another dimension of space. The idea goes back to the 1980s, when Hawking and Hartle sought to explain the big bang as the moment when time and space became differentiated. In 2008 Marc Mars of the University of Salamanca in Spain and José M. M. Senovilla and Raúl Vera, both at the University of the Basque Country in Spain, applied a similar idea not to time's beginning but to its end.

They were inspired by string theory and its conjecture that our four-dimensional universe—three dimensions of space, one of time—might be a membrane, or simply a "brane," floating in

© 2018 Scientific American

a higher-dimensional space like a leaf in the wind. We are trapped on the brane like a caterpillar clinging to the leaf. Ordinarily, we are free to roam around our 4-D prison. But if the brane is blown around fiercely enough, all we can do is hold on for dear life; we can no longer move. Specifically, we would have to go faster than the speed of light to make any headway moving along the brane, and we cannot do that. All processes involve some type of movement, so they all grind to a halt.

Seen from the outside, the time lines formed by successive moments in our lives do not end but merely get bent so that they are lines through space instead. The brane would still be 4-D, but all four dimensions would be space. Mars says that objects "are forced by the brane to move at speeds closer and closer to the speed of light, until eventually the trajectories tilt so much that they are in fact superluminal and there is no time. The key point is that they may be perfectly unaware that this is happening to them."

Because all our clocks would slow down and stop, too, we would have no way to tell that time was morphing into space. All we would see is that objects such as galaxies seemed to be speeding up. Eerily, that is exactly what astronomers really do see and usually attribute to some unknown kind of "dark energy." Could the acceleration instead be the swan song of time?

YOUR TIME IS UP

BY THIS LATE STAGE, it might appear that time has faded to nothingness. Yet a shadow of time still lingers. Even if you cannot define duration or causal relations, you can still label events by the time they occurred and lay them out on a time line. Several groups of string theorists have recently made progress on how time might be stripped of this last remaining feature. Emil J. Martinec and Savdeep S. Sethi, both at the University of Chicago, and Daniel Robbins of Texas A&M University, as well as Horowitz, Eva Silverstein of Stanford University and Albion Lawrence of Brandeis University, among others, have studied what happens to time at black hole singularities using one of the most powerful ideas of string theory, known as the holographic principle.

A hologram is a special type of image that evokes a sense of depth. Though flat, the hologram is patterned to make it look as though a solid object is floating in front of you in 3-D space. The holographic principle holds that our entire universe is like a holographic projection. A complex system of interacting quantum particles can evoke a sense of depth—that is to say, a spatial dimension that does not exist in the original system.

But the converse is not true. Not every image is a hologram; it must be patterned in just the right way. If you scratch a hologram, you spoil the illusion. Likewise, not every particle system gives rise to a universe like ours; the system must be patterned just so. If the system initially lacks the necessary regularities and then develops them, the spatial dimension pops into existence. If the system reverts to disorder, the dimension disappears whence it came.

Imagine, then, the collapse of a star to a black hole. The star looks 3-D to us but corresponds to a pattern in some 2-D particle system. As its gravity intensifies, the corresponding planar system jiggles with increasing fervor. When a singularity forms, order breaks down completely. The process is analogous to the melting of an ice cube: the water molecules go from a regular crystalline arrangement to the disordered jumble of a liquid. So the third dimension literally melts away.

As it goes, so does time. If you fall into a black hole, the time on your watch depends on your distance from the center of the hole, which is defined within the melting spatial dimension. As that dimension disintegrates, your watch starts to spin uncontrollably, and it becomes impossible to say that events occur at specific times or objects reside in specific places. "The conventional geometric notion of spacetime has ended," Martinec says.

What that means in practice is that space and time no longer give structure to the world. If you try to measure objects' positions, you find that they appear to reside in more than one place. Spatial separation means nothing to them; they jump from one place to another without crossing the intervening distance. In fact, that is how the imprint of a hapless astronaut who passes the black hole's point of no return, its event horizon, can get back out. "If space and time do not exist near a singularity, the event horizon is no longer well defined," Horowitz says.

In other words, string theory does not just smear out the putative singularity, replacing the errant point with something more palatable while leaving the rest of the universe much the same. Instead it reveals a broader breakdown of the concepts of space and time, the effects of which persist far from the singularity itself. To be sure, the theory still requires a primal notion of time in the particle system. Scientists are still trying to develop a notion of dynamics that does not presuppose time at all. Until then, time clings stubbornly to life. It is so deeply engrained in physics that scientists have yet to imagine its final and total disappearance.

Science comprehends the incomprehensible by breaking it down, by showing that a daunting journey is nothing more than a succession of small steps. So it is with the end of time. And in thinking about time, we come to a better appreciation of our own place in the universe as mortal creatures. The features that time will progressively lose are prerequisites of our existence. We need time to be unidirectional for us to develop and evolve; we need a notion of duration and scale to be able to form complex structures; we need causal ordering for processes to be able to unfold; we need spatial separation so that our bodies can create a little pocket of order in the world. As these qualities melt away, so does our ability to survive. The end of time may be something we can imagine, but no one will ever experience it directly, any more than we can be conscious at the moment of our own death.

As our distant descendants approach time's end, they will need to struggle for survival in an increasingly hostile universe, and their exertions will only hasten the inevitable. After all, we are not passive victims of time's demise; we are perpetrators. As we live, we convert energy to waste heat and contribute to the degeneration of the universe. Time must die that we may live. SA

MORE TO EXPLORE

Toward the End of Time. Emil J. Martinec, Daniel Robbins and Savdeep Sethi in *Journal of High Energy Physics*, Vol. 2006, No. 8; August 16, 2006. http://arxiv.org/abs/hep-th/0603104

Is the Accelerated Expansion Evidence of a Forthcoming Change of Signature on the Brane? Marc Mars, José M. M. Senovilla and Raúl Vera in *Physical Review D*, Vol. 77, No. 2; January 11, 2008. http://arxiv.org/abs/0710.0820

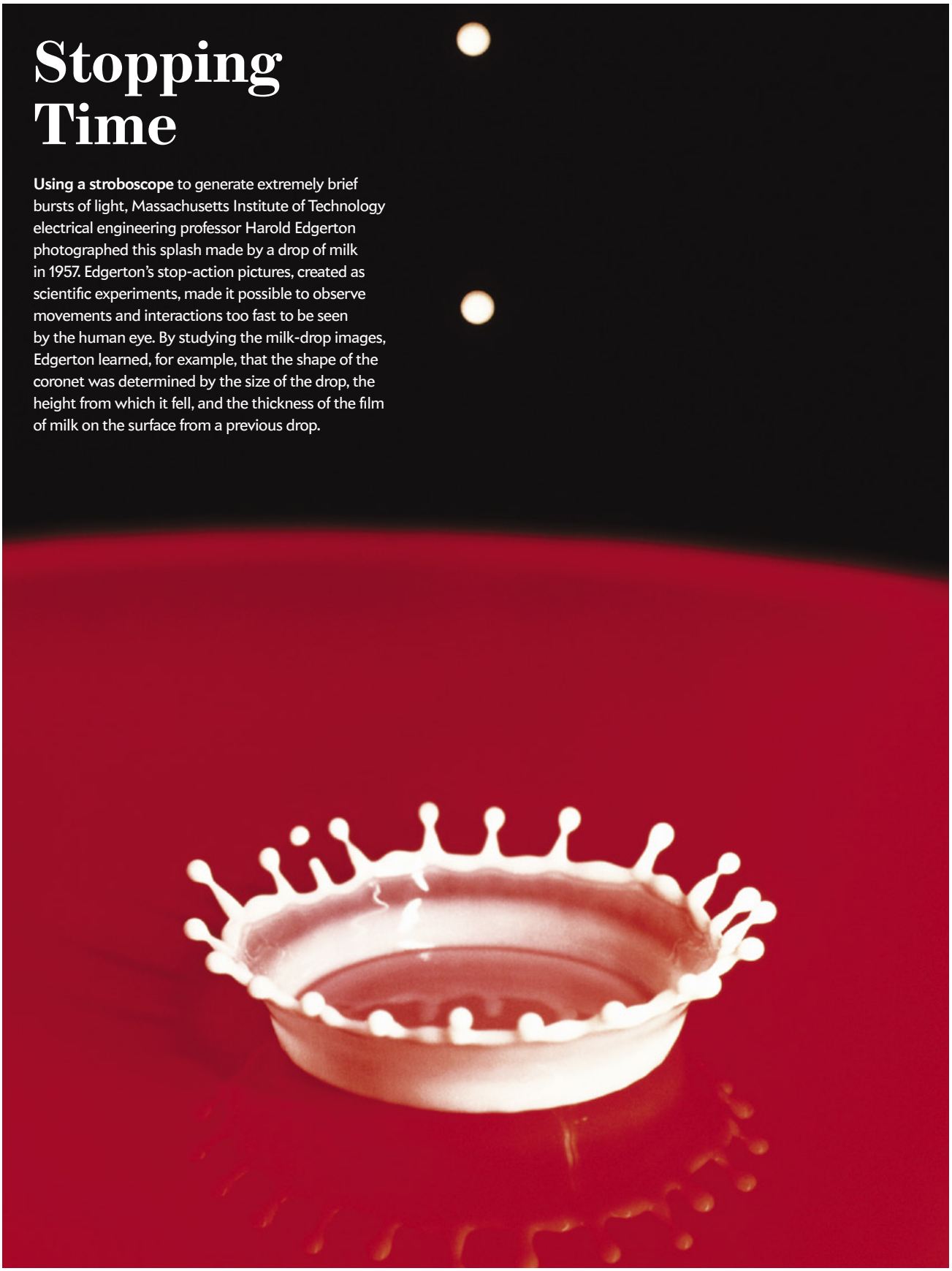
Cycles of Time: An Extraordinary New View of the Universe. Roger Penrose. Knopf, 2011.

scientificamerican.com/magazine/sa

© 2018 Scientific American

Stopping Time

Using a stroboscope to generate extremely brief bursts of light, Massachusetts Institute of Technology electrical engineering professor Harold Edgerton photographed this splash made by a drop of milk in 1957. Edgerton's stop-action pictures, created as scientific experiments, made it possible to observe movements and interactions too fast to be seen by the human eye. By studying the milk-drop images, Edgerton learned, for example, that the shape of the coronet was determined by the size of the drop, the height from which it fell, and the thickness of the film of milk on the surface from a previous drop.



HAROLD EDGERTON *Science Source*

© 2018 Scientific American

BRIGHT HORIZONS 36

SCIENTIFIC AMERICAN Travel

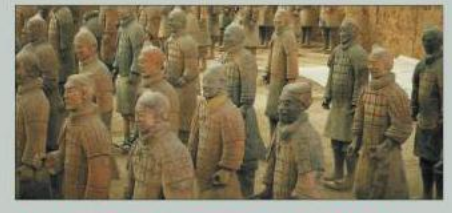
China | October 27th – November 6th, 2018

Embark on Bright Horizons 36's journey of discovery in China and send your senses reeling. Taste dim sum in Hong Kong and local brewed beer in Qingdao. Soak in powder blue hot springs and taste the umami of oyster pancakes in the Taiwanese night markets. Hear the roar of urban life and share the quiet of China's temples and gardens with a friend.

While we sail the Yellow Sea and the East and South China Seas, absorb Earth science from an expert and study the roots of Chinese culture in exclusive Bright Horizons classes.

The time is ripe to visit China and Taiwan. Update your Earth science knowledge and get an unparalleled view of China and her history and culture.

Cruise prices vary from $3,599 for an Ocean View Stateroom to $9,624 for a Owner's Suite, per person (pp) based on double occupancy. For those attending our SEMINARS, there is a $1,675 fee. Add'l pp fees: gov't taxes and fees ($440), a non-refundable Booking Service Fee ($150), and the Tour Leader gratuities ($100), and the option Program/Conference fee ($1,675). The Program, cruise pricing, and options are subject to change. For more information email us at info@InsightCruises.com.



ULTIMATE CHINA

PRE-CRUISE LAND TRIP OCTOBER 19TH – 27TH, 2018

Enhance your once-in-a lifetime trip to China with an optional 8-night, 5-star precruise tour of Beijing, Xi'an's Terracotta Warriors, and rural Guilin's iconic landscapes with Insight Cruises and Imperial Tours. You'll enjoy privileged access behind the scenes in the Forbidden City's "Palace of Double Glory"; see the Terracotta Warriors as visiting dignitaries do, walking between and among them; view 1,000 y.o. Tang frescos up close. Dine *on* the Great Wall, raft through rural China, visit a distinguished art collector — immerse yourself in China's ways to enrich your Bright Horizons cruise conference.

If you have wanted to see China's key sites but hesitate because of language, unfamiliarity with available housing, intra-China travel arrangements (this package includes three flights), or analysis paralysis, we can take you there. Effortlessly. Accompanied by a China host as well as very select area guides, you will see China's touchstones in a seamless and unforgettable experience.

Insight Cruises and Imperial Tours will take care of all the details so you can focus on China with peace of mind.

SEMINARS

EARTH SCIENCE:

- Scales of Change: Why Time Matters to Science and Being Human
- Climate Through the Ages
- How Mountain Building and Climate Shape the Land
- The Early Earth and Life's Origins
- Life: A Brief History
- The "Ice Ages", Rapid Climate Change, and Human Impacts

SPEAKER: Ron Martin, Ph.D.

UNDERSTANDING CONTEMPORARY CHINA:

- How History Informs Contemporary Chinese Attitudes and Actions
- China as a Global Power

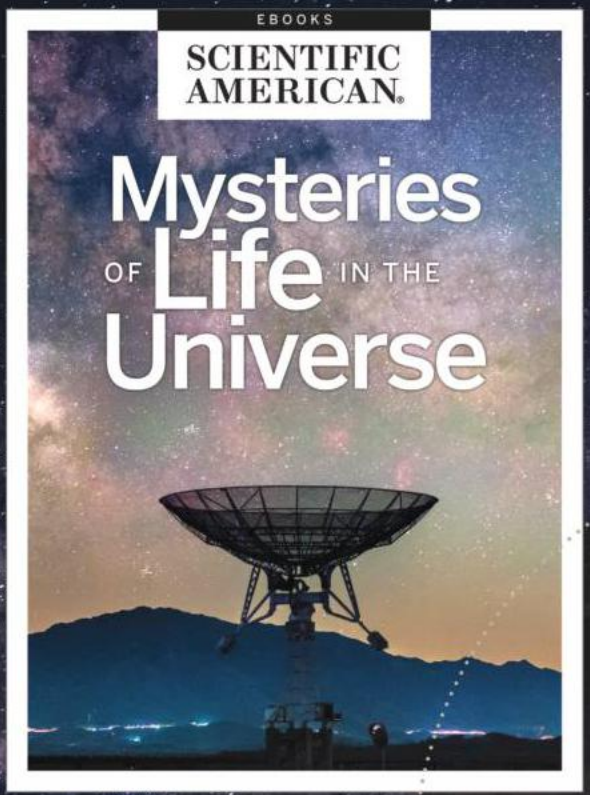
EXPERIENCING IMPERIAL CHINA:

- The Grand Canal and the Great Wall
- Being a Chinese Peasant
- Being a Chinese Scholar
- Imperial China and the Maritime World

SPEAKER: Andrew Wilson, Ph.D.

FOR MORE INFO Please email: **info@InsightCruises.com** or visit: **ScientificAmerican.com/travel**

CST# 2060380-40 ©Scientific American, Inc.



Examine the Big Questions

How did life begin?
Are we alone in the universe?
In this eBook, we explore several theories on the origin of life, some of its extreme forms and the ongoing search for signs – or sentience – on distant worlds.

sciam.com/ebooks

SCIENTIFIC AMERICAN eBooks